(2)

Research Note 85-20

AD-A160 418

System Development and Evaluation Technology: State of the Art of Manned System Measurement

Joan M. Edwards, Richard F. Bloom, and Paul A. Brainin
Dunlap and Associates, East, Incorporated

ARI Field Unit at Fort Benning, Georgia
Seward Smith, Chief

Training Research Laboratory
Harold F. O'Neil, Jr., Director

U. S. Army
Research Institute for the Behavioral and Social Sciences

February 1985

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

L. NEALE COSBY
Colonel, IN
Commander

EDGAR M. JOHNSON
Technical Director

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 85-20	2. GOVT ACCESSION NO. AD-A160418	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) System Development and Evaluation Technology: State of the Art of Manned System Measurement		5. TYPE OF REPORT & PERIOD COVERED Final Report 1980 - 1982
		6. PERFORMING ORG. REPORT NUMBER 293-8
7. AUTHOR(s) Joan M. Edwards, Richard F. Bloom, and Paul A. Brainin		8. CONTRACT OR GRANT NUMBER(s) MDA903-80-C-0345
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dunlap and Associates East, Inc. One Parkland Drive Darien, Connecticut 06820		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263743A794
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE February 1985
		13. NUMBER OF PAGES 85
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) --		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

Seward Smith, Contracting Officer's Representative

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Manned System, Taxonomy, Model, Evaluation, Performance Measurement, Effectiveness Measurement, State of the Art

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report is an up-to-date assessment of the state of the art of manned system measurement. The assessment is based in part on the material presented in the Task 3a report-Review of Manned Systems Measurement Literature.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

This report is an up-to-date assessment of the state of the art of manned system measurement. The assessment is based in part on the material presented in the Task 3a report—Review of Manned Systems Measurement Literature. It reflects the review and abstracting of over 250 relevant technical documents.

This report employs a topic outline compatible with the overall measurement model being developed under the present contract. ~~Nevertheless, it is believed~~ that the model is sufficiently representative and comprehensive so that all significant comments and authors have a place in its structure.

One of the important uses of this report is the identification of current measurement capabilities and limitations, so that requirements and priorities for the improvement of system-oriented measurement can be delineated. In this review, it became apparent, for example, that measurement models need to be further developed, supported with appropriate human performance data, refined through more consistent and comprehensive applications, and validated by independent corroborations of some kind. Furthermore, the general sense of impracticality, and the need for simplifying assumptions in some cases, strongly suggests a requirement for improving the "efficiency" of measurement models by reducing the magnitude of effort required in their application. It is envisioned that much time, effort, and money can be saved, irrelevant measurements can be avoided, meaningfulness and utility can be enhanced, and additional applications of the models can be found if several key improvements are made.

TABLE OF CONTENTS

I. INTRODUCTION

This report was prepared under the overall contract for the "Study of Effectiveness of Infantry Systems: TEA, CTEA, and Human Factors in Systems Development and Fielding" (MDA903-80-C-0345). Dunlap and Associates, Inc., is responsible for Task 3 (System Development and Evaluation Technology) of that contract, under subcontract (No. 05628) to the Mellonics Systems Development Division of Litton Systems, Inc. The present report is in partial fulfillment of Task 3c, "Analyze and Synthesize the Results." Tasks 3a and 3b of the Dunlap effort involve a literature search in the area of manned systems measurement, and further development of the Systems Taxonomy Model (STM), respectively.

The principal end product of Task 3 will be a model for the overall process of measuring the performance and effectiveness of manned systems. It is not expected that this will be a fully developed overall process model; it is highly likely that such full development will require research that is beyond the present scope of work. However, it is expected that this task will accomplish a good deal of the initial development that is required, will advance the measurement state of the art, and will produce the solid foundation for the future full development of the overall process model.

The present report uses the Task 3a abstracts of over 250 documents as a point of departure to compile and present an up-to-date assessment of the state of the art of manned system measurement. It addressed measurement limitations as well as capabilities, so that requirements and priorities for improvement can be clearly delineated. Particular attention is paid in this evaluation to the issue of system-oriented measurement. A "system" is taken to include people, equipment and operating procedures.

II. MANNED SYSTEM MEASUREMENT: GENERAL

The identification and acquisition of relevant manned system measurement literature was built on an existing base of documentation. This base consisted of the searches conducted by ARI of the NTIS and DDC (now DTIC) data bases in Feburary 1977. The ARI literature file was updated and extended by conducting searches using the same data bases and key words to acquire new entries since the original search was performed. In addition to the NTIS and DTIC searches, the present search was expanded to include the PASAR and COMPENDEX data bases.

Using the literature search results as a partial guide, a framework was developed for the purpose of enumerating (at least in general terms) the steps in an overall conceptual process model for measuring the performance of effectiveness of any human-machine system. This enumeration was used for structuring the review/annotation of relevant literature during Task 3a, and is used in a similar way for this report. The steps are illustrated in Figure 1 and can be described briefly as follows:

1. Definition of the System

   At the outset of the measurement process, the analyst must determine with what kind of system he or she is dealing.

2. Definition of the System's Missions

   The analyst needs to know exactly what kinds of job this system is supposed to perform. Ultimately, it is the system's ability to do those jobs that will determine how well the system performs.

3. Specification of the Environment

   Performance measurement ultimately must reflect how well the system will do its jobs under realistic circumstances. Thus, the analyst needs to know where, when, and under what conditions those jobs need to be done.

4. Specification of the General Constraints

   The analyst needs to know all of the limitations and conditions that will be imposed on the system and its jobs so that fully realistic measurement can occur, and so that all relevant issues can be examined.

5. Identification of the Ultimate Performance Requirements

   The analyst needs to determine, in general terms, exactly what outputs, products, or end results are supposed to occur from the successful performance of the system's jobs.

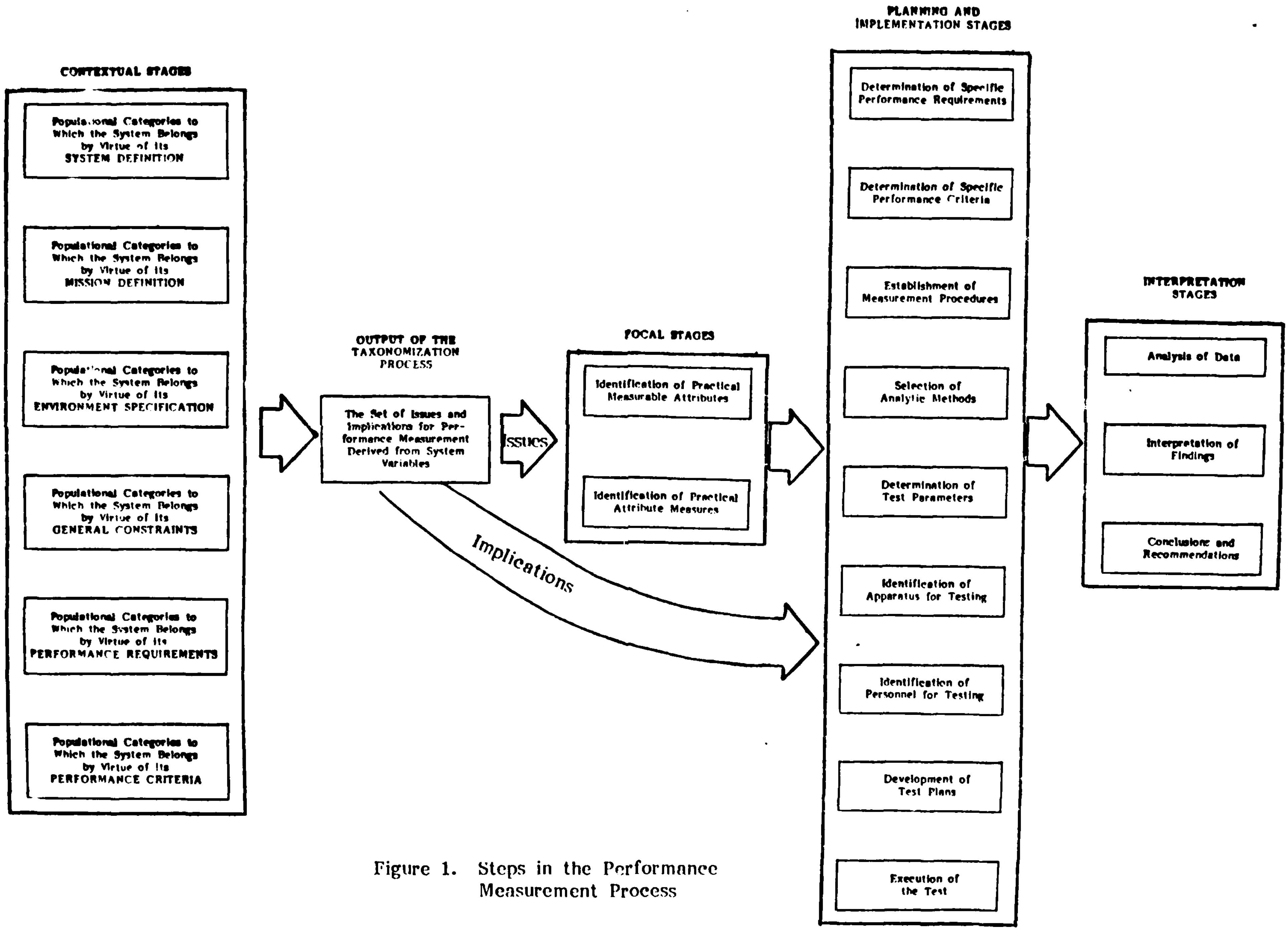

Figure 1. Steps in the Performance Measurement Process

6. Identification of the Ultimate Performance Criteria

> The analyst needs to know, again in general terms, how to determine whether these outputs, products, or end results are adequate.

It should be noted that these six stages constitute the "what, when, and where" of system performance measurement. Once he or she has completed the sixth step, the analyst will know what the system is, what jobs it has, what results it seeks to achieve, and where, when, and under what limitations it is supposed to operate. These, then, are the contextual stages of the measurement process, and also constitute what has been termed the System Taxonomy Model (STM) in this project.

The enumeration of steps in the overall conceptual process model continues:

7. Identification of Practical Measurable Attributes

> Once the analyst knows what the system is supposed to do and what results it is supposed to produce, he or she must identify concrete, observable events, effects, and phenomena that can be used to determine whether or not the jobs have been done and the results produced. These might be events, effects, or phenomena that themselves stem directly from the system's performance of its job. Alternatively, they might stem from the failure of the job, or be associated in some indirect way with the system's *accomplishments* or failures.

8. Identification of Practical Attribute Measures

> Having identified the events, effects, and phenomena that can help to determine whether or not the system has done its job, the analyst needs to choose some means of handling those outcomes to assess how much of the job has been accomplished and how well it has been performed. This entails the application of some "yardsticks" or computations to the attributes chosen as indicators of performance. That is, if the attribute of interest is some phenomenon, the measure might be how often the phenomenon occurs, how long it lasts, or how large it is. The measure might also involve some comparative computation involving that phenomenon and some other, undesirable phenomena, such as a ratio between "good" and "bad" effects.

It should be noted that the preceding two stages constitute another important milestone in the overall process of human machine system measurement. They might be termed the focal stages of the process, in the sense that the measurable attributes and the attribute measures are the things on which the analyst focuses when he or she conducts the assessment of system performance.

The listing of overall measurement process steps continues:

9. Identification of Specific Performance Requirements

At this point, the analyst needs to translate the general expression of the system's intended outputs, products, or end results into terms specifically keyed to the selected measures.

10. Identification of Specific Performance Criteria

Similarly, the general expression concerning how to determine whether the system's outputs are adequate must be translated into measures-specific terms.

11. Specification of Measurement Procedures

At this point in the process, the analyst begins to specify the technical and procedural details concerning the measurement application at hand. The first concern is with the procedures for generating the selected measures, including specification of the data that are needed, where and when these data can be collected, how to collect the data, how to insure quality control over the collection process, and other related concerns.

12. Specification of Analytic Methods

Before the data are collected, the analyst must determine exactly what he or she will do with those data. The statistical tests to be employed, the combinatorial procedures to be used, and the level of precision desired all will affect the scope of the measurement application (such as the sample size) and the kinds of conclusions that can be reached.

13. Determination of the Test Parameters

The analyst must decide which conditions will be varied, which will be held fixed, how data will be grouped into class intervals, how many measurement replications will be conducted, and the various other parameters associated with application of the selected measures and the selected analytic methods.

14. Determination of the Apparatus Needed for Testing

The analyst must specify what equipment will be used in the measurement process, the format of the data that the equipment will produce, any format or data media changes that may be needed, and similar equipment-related issues.

15. Determination of the Personnel Needed for Testing

This concerns both the personnel who will conduct the test (as data collectors, analysts, administrators, logistic support, etc.) as well as the people who will operate the system during testing (test subjects). In each case the analyst must specify the numbers

of people needed, the qualifications they must have, relevant demographic or other characteristics which they must have, the pre-test training they are to receive, and other relevant factors.

16. Preparation of the Test Plan

This is a summarizing step in the process, during which the analyst's decisions during the preceding seven steps are formally documented for review, reconsideration and revision, and finally implemented.

17. Execution of the Test

Ultimately, the analyst puts the test plan into operation by conducting the test and applying the measures in accordance with the procedures selected in the previous steps.

These last nine steps constitute what may be termed the planning and implementation stages, during which the measures that emerge from consideration of the system as a member of many population categories are applied to assessment of the system's performance. These are by no means trivial steps. If they are conducted without skill or care, the effort that went into selection of the measures may be wasted, and a misleading assessment of the system's performance may be produced. However, while never denying the importance of these planning and implementation stages, one should bear in mind that the outcome of those stages can (at best) be only as good as the measures that were chosen. If the measures set includes some that are inappropriate and/or misses some that are highly pertinent, an improper assessment of system performance likely will result no matter how carefully the test is planned and executed.

The enumeration of measurement process steps concludes with the following three:

18. Analysis of Data

In accordance with the methods and techniques previously selected, the analyst must combine and manipulate the data to generate the measures and produce the quantitative and qualitative bases for assessing system performance.

19. Interpretation of Findings

Using statistical and other appropriate techniques, the analyst must examine the measures and combinations of measures and determine how much and how well the system has done its jobs.

20. Development of Conclusions and Recommendations

Finally, the analyst must apply the findings to the original measurement purposes and answer the questions that motivated the measurement effort. These might include such questions as:

> Is the system *feasible*? Is it *cost effective*? Is it, overall, better or worse than some other, competing system? Should its development be continued? What design changes are needed?

The last three steps may be termed the interpretation stages of the process. They represent the final outcome of performance measurement and its application to the particular research issues at hand.

These 20 steps represent only one conceptual formation of the total measurement process in terms of its constituent activities. Other analysts might use different terminology to describe the processes' stages, and might identify more or fewer activities depending on how finely-grained a view they wish to take. However, the authors believe that most analysts would agree that these 20 steps provide a fair and valid representation that completely covers the system measurement process from start to finish, thereby serving as a convenient structure for this state of the art review.

III. CURRENT MEASUREMENT CAPABILITIES

A. General

Measurement of performance and effectiveness has been going on for a long time, and many pure and applied research efforts for assessing manned systems capabilities and limitations have been reported. Widely accepted and frequently used analytic techniques abound. The most relevant prior work on manned systems measurement and associated taxonomies to help define and facilitate implementation was done by Finley and her colleagues (1975, 1976). From their work on Systems Measurement Theory, and on System Taxonomy Models (STMs) specifically, it can be seem that certain prerequisites exist for including "system" factors in manned system performance measurement. They are:

- Recognition of systems as viable entities in and of themselves.
- Development of conceptual tools for the purpose of:
  - Grouping systems into populations.
  - Defining these populations.
  - Placing them into a context with other populations.

In all cases, the basic purpose of a taxonomy is to supply knowledge that is specifically relevant to the particular analytic application at hand. Thus, each system taxonomy is unique to the particular system and to the particular context and purposes in and for which the measurement process is to be applied. What Finley et al. are seeking is a systematic way of generating such taxonomies for any given system and measurement purpose. Development of the STM by those researchers and in the current project is intended to help meet that need. Within an overall conceptual process model for evaluation, the STM is a tool that will support the taxonomy development process for manned system studies. Its purpose is to aid the analyst in developing conceptualizations of:

- Systems as entities which form populations
- Populations taxonomies, including both system populations and system aspect populations (e.g., its missions, performance requirements, etc.)
- System taxonomies, i.e., organizations of any given system's populations class and distinguishing characteristics that are relevant to measurement research dealing with that system

Initial development of the STM by Finley et al. focused on three concepts:

1. Measurement Level Definitions

The two general measurement levels are nominal and relative. The relative level includes the ordinal, interval, and ratio categories familiar from elementary statistics. Measurement levels are relevant to the STM basically because the taxonomies sought here can be viewed as sets of measures and measure relationships.

2. Levels of System Description

The three levels of system description are: 1) system objectives, 2) system functional purposes, and 3) the various system activities, characteristics, and requirements. There is a correspondence between these system description levels and the measurement levels: system objectives tend to generate families of nominal measures, while system activities, characteristics, and requirements produce relative measures; the system functional purposes can produce both nominal and relative measures.

3. Types of Questions

The research questions or issues which the analyst faces are many and varied, but generally fall into two types: fundamental research vs. applied research. The type of question will affect which Levels of System Description are appropriate for generating taxonomies suited to the measurement application at hand.

The STM is the general form within which all particular system taxonomies would fit. Prior to this project, the model had been carried out to a preliminary stage of development, which is depicted in Figure 2. The work under the present contract made use of that model and the following concepts as points of departure.

- The operator or crew of any "manned" system must be viewed as one of several system elements, along with equipment and operating procedures.

- "Manned" systems are viable entities in and of themselves, and often can be grouped in a context with other systems to form definable populations.

- The STM is intended to help insure that all system elements (people, equipment and procedures) are incorporated into the process of generating performance measures.

- The general STM is applied uniquely to a particular system, in a particular context, to satisfy a particular measurement purpose.

- The specific system taxonomies developed using the STM can be viewed (in the abstract) as sets of measures and measure relationships, those measures including any of the nominal, ordinal, interval and ratio categories.

- Systems can be described at various levels of generality or detail, each of which can generate required measures for the particular purpose at hand. The STM can help the analyst to keep all levels of system functioning in mind, thereby increasing the likelihood of generating a complete and efficient set of measures.

- The STM is part of an overall model for the entire process of measuring the performance of "manned" systems, and therefore must be designed to be compatible with that larger, overall model.

	MEASUREMENT LEVELS	SYSTEM TAXONOMIC LEVELS	EXAMPLES OF POSSIBLE TAXONOMIC CATEGORIES & DIMENSIONS
LEVEL ONE	Nominal Measurement	SYSTEM OBJECTIVES	• Production • Supply • Navigation • Air Traffic Control • Health & Welfare • Transportation • Maintenance • Weapons • Surveillance • Etc.
LEVEL TWO	Nominal ⇑	SYSTEM FUNCTIONAL PURPOSES	Nominal • Indirect command/control/guidance operations • Relatively direct control/navigation operations • Maintenance operations • Data or materials processing
	⇓ Relative		Relative • Command • Control • Information • Data
LEVEL THREE	Relative Measurement (Ordinal, Interval and Ratio)	STRUCTURAL CHARACTERISTICS	• Organization and layout • Size • Level of automation • Implementation capabilities
		OPERATOR/EQUIPMENT CHARACTERISTICS	• Human skills, equipment conditions • Human abilities & IQs, equipment capabilities • Values • Needs
		OPERATING CHARACTERISTICS	• Inputs to operator • Operator processing • Operator outputs • Units being dealt with by system • Environment • Feedback
		SUPPORT REQUIREMENTS CHARACTERISTICS	• Materials (including people) • Maintenance (including people)

From: Finley & Muckler (1976) and Finley et al. (1975)

Figure 2. A General Systems Taxonomy Model (Initial Status)

Forthcoming reports under the balance of the present study will present more detailed and advanced thinking in this area. The remainder of this section reports the state of the art by other researchers.

B. State of the Art Review

The following paragraphs describe what other researchers are currently doing in the area of systems measurement and analysis. For convenience and compatibility with the work previously conducted on this project, the state of the art is reviewed in terms of those topical areas defined in Section II.

1. General System Measurement

Most of the reports covered in this review had something to say about general systems measurement (James, 1972; Knoop, 1978; Rouse,1977; Quinn, 1970; Cogan et al. 1972; and Siegel et al. 1974, for example). It would be too laborious and of little use to the reader to present all of these thoughts here. Instead, this section contains a summary of the most significant statements concerning this topic with an attempt to keep redundancy to a minimum.

The report by Markel (1965) considered the issues involved in a general theory of systems evaluations. It discussed the need to define the evaluation problems in a particular system, and the need to break down the problem into smaller more tractable problems to facilitate a workable approach to any evaluation. The underlying substance of the evaluation process is measurement, and the key to successful measurement and evaluation is to be found in criteria selection. Further, it was suggested that the development of a general theory of systems evaluation can be approached by identifying and defining those elements which can provide a basis for overall evaluation of any system. These elements were described in three broad areas of primary concern for systems in general: systems structure, systems operation, and systems performance.

According to Mitchell et al. (June 1967), allocation of system effectiveness requirements is the process of determining how the total system's effectiveness requirements distribute among the system's constituent man-machine functional units/states. To develop a procedure for effectiveness requirements allocation, guidelines can be generated for: 1) specifying the system effectiveness requirements along all its dimensions; 2) partitioning the system into requirements and states; 3) characterizing and specifying input data; and 4) relating the system's effectiveness requirements to system segments consistent with the input data. Chop (1972) stated that system effectiveness is based on a quantitative measure of the extent to which the system is expected to meet its assigned role in a specific mission. The measure is dependent upon system parameters of availability, dependability and capability. Sheldon et al. (1967) felt that it is increasingly evident that man-machine system evaluation needs techniques that are radically different from traditional methods. The overall purpose of their model is to develop a methodology permitting evaluation of man-machine performance based on a series of flexible standards reflecting the difficulty of the mission, in direct contradistinction to the absolute standards approach.

Baker (1970) developed a general information system model which focused on man and considered the computer as a tool. The ultimate objective was to produce a simulator which would yield measures of system performance under different

mixes of equipment, personnel and procedures. Among the immediate benefits of the model is the potential to quantify human performance by employing system measures.

In man-machine systems, according to Connelly et al. (1976), the human operator adapts his control characteristics so that the overall system response satisfies his performance criteria. The system designer should have available a design tool that provides a means for estimating the operator's performance criteria and his control actions, so that the designer can determine which design features support performance and which features degrade performance. Another report (Geddie, 1976) discussed methods to control the variance contributed by the human operator which influences the total system performance.

As an example of the development of system measurement, the report by Jahns (1973) represents an initial attempt, through a literature review, to scope the complexity of developing a conceptual structure of operator workload in the operation of a vehicle system. The ultimate goal was to develop a quantitative index of operation performance for any point in time during operation. Buckley et al. (1976) described a performance measurement system for air traffic controllers.

Another paper (Siegel, 1978) discussed the methods for measuring human performance reliability and methods for integrating human performance reliability with equipment reliability to derive a measure of total system reliability. Emphasis was placed on a computer simulation model that was basically a sequential processor which incorporated human, equipment and mission factors. The evaluation factors were mission effectiveness, time utilization, personnel, and report frequency. Meister (1968) pointed out that any measurement of system reliability or system effectiveness which does not include indices of human performance must necessarily produce an erroneous estimate of that system's reliability of effectiveness.

On a higher system level, four papers were concerned with evaluation. The first (Churchman, 1971) said that organizations are goal oriented and the goal structure can be translated into measures of performance such as profitability, benefit minus cost, social utility, etc. The second study (DiGialleonardo et al. October 1974) provided a model for assessing the benefits and costs in management and information systems, while the third (Connelly et al. October 1969) provided a workable cost/effectiveness methodology for man-machine function allocation. The fourth paper (Willis, 1967) provided a methodology which enables cognizant persons to obtain quantitative information on personnel effectiveness and relative costs.

Operational testing or evaluation is a general form of measurement of a system. It is often discussed as it is utilized in the military. According to McKendry et al. (1964), an operational evaluation is the test and analysis of a weapon system, support system, component or equipment under service operation conditions, insofar as practical, to determine the ability of a system, component, or equipment to meet specified operational performance requirements and/or to establish suitability for service use. An operation test was defined by Montgomery et al. (1975) as that test and evaluation conducted to estimate the prospective system's utility, operational effectiveness, and operational suitability. One of the objectives of operational testing is an independent evaluation of competing systems resulting in some statement of relative attributes and preferences.

Williams (1975) developed a method used in operational testing which provides a basis for the selection of critical attributes which best discriminate between acceptable and unacceptable systems. According to Analytics, Inc. (1976), the test methodology must assess the functional performance of the system and cannot be designed to match the system itself, lest the result be determined by the evaluation method. The purpose of another study about operational testing in the military (Rankin, 1975) determined the use of fault free analysis in operational test planning.

The concept of "system" is the formulation of human factors studies since the latter seek to measure factors that affect personnel performance in manned systems (Meister, 1978). An important aspect of system performance to be considered is the field of human factors, that is, human performance in relation to system performance (McKendry et al. 1964). Systems measurement is a means of focusing step-by-step on the human performance aspects of the system to be enhanced and identifying the interrelationships of the human factor system variables in order to determine productivity under varying condition (Uhlaner, 1970).

It has been stated in the military, according to Miles (1976), that the soldier is part of the system and human factors data should be analyzed, not as a separate additional activity, but as an integral part of the evaluation of each system. Human engineering services and end products relating to assessment of system performance include (Coburn, 1973): 1) man-machine concept analyses—prediction of man-related aspects of system performance for candidate or selected system configurations; 2) man-machine system design—establishment of performance specifications which set bounds on man-machine system performance and define what the system must do in operational terms.

Over the past three decades, there has been an increasing demand for quantitative techniques of human performance prediction in man-machine systems tasks. A somewhat bewildering variety of methods has evolved to satisfy this need, ranging from specific task simulation to classical tests of fundamental human abilities (Finley et al. 1970). The Technique for Establishing Personnel Performance Standards TEPPS) is designed as a performance tool (Smith et al. 1969, Vols. 1 and 2; and Mitchell et al. August 1967). TEPPS has two primary objectives: 1) deriving specific personnel performance standards with definable relations to system effectiveness requirements; 2) determining the influence on system effectiveness of performance levels that deviate from established performance standards. The HRTES is a systematic and integrative approach to planning and conducting evaluation of human contributions to system performance (Kaplan et al. 1978). It encompasses a set of procedures which will assure that human resources are properly included in system design and are adequately assessed and evaluated during operational test and evaluation. Other techniques for performance measurement are described by Uhlaner et al. (1980). These include: Skill Qualification Test (SQT), Organizational Effectiveness (OE) programs, Work Environmental Questionnaire (WEQ), the System Measurement Bed (SMB), and others. In addition, there is the TART (Task Analysis Reduction Technique) which allows for the facilitation of human performance quantification, clarification of analysis and improved usability of the data (Ellis, 1970).

The military has sponsored numerous projects concerning the measurement of weapon systems. The studies conducted by Larson et al. (1974), and Gex (1961) contains literature surveys relevant to this. The Weapon System Effectiveness Industry Advisory Committee (1965, Vols. 1 and 3)—as an example of one group

in this area—had as one responsibility to the Air Force Systems Command to recommend uniform methods and procedures to be applied in predicting and measuring systems effectiveness during all phases of a weapon system program.

Various other reports describe more specific weapon systems. Klein (undated) and Klein et al. (1969) discuss the development of combat related measures and operational test procedures for small arms weapon system evaluation. Rankine (1970) and Burgin et al. (1972) describe measurement of aircraft systems and performance. Sonar (Fischl et al. 1968), radar (Sidoruk, 1977) and ordnance systems (Lindsey, 1974) measurement are other examples.

Training system evaluation programs and techniques of measurement were topics in several reports reviewed (e.g., Bond et al. 1970; Lyons, 1972; U.S. Department of the Army, 1975; Hansen et al. 1974; Sjogren, undated; Hammell et al. 1973; Ford et al. 1974; and Dieterly, 1973). According to Narva (1978), the development of the training subsystem must occur concurrently with that of the prime system in order to meet the objectives of having a total system operational when fielded. The goal of operational testing in general is to identify a general learning curve which can be described as a mathematical function. This type of formulation would enable measurement of the impact of the training level of a crew or unit engaged in operational tests (Brokenburr, 1978).

2. Systems Taxonomy Model

The Systems Taxonomy Model is intended to serve as a tool in the improvement of manned systems measurement of performance and effectiveness. It is hoped that such a model will enable researchers to include the appropriate system design and operational factors in their studies. In addition to recognition of systems as viable entities, the Systems Taxonomy Model will provide the conceptual tools for grouping, defining and placing system populations into a context with other populations. Taxonomization is the process of first collecting together the relevant variables, factors and characteristics of that system and, second, finding some identification and organization of those things which will make them more manageable, tractable, or simply more understandable.

Of the literature reviewed regarding the specific development of a systems taxonomy model, the most extensive discussions are to be found in Finley et al. (1970, 1975, 1976)—studies which have been reviewed and cited earlier in this chapter. Several other researchers have attempted to formulate classification schemes in certain areas of the system measurement process, and other have theorized on the need for such a taxonomy and how it might be developed.

In a study conducted by Tien (1979), in what the author termed only an initial step toward a systematic approach to program evaluation design, an attempt was made to synthesize and systematize the steps necessary to develop valid and comprehensive evaluation designs. In the first step, a design framework is identified which links program characteristics to design elements through an expanded set of threats to validity. Secondly, the various design elements are grouped into five systematically convenient components including test hypothesis, selection scheme, measures framework, measurement methods and analytic techniques. Thirdly, it was proposed that different types of evaluation can be contained in an evaluation taxonomy composed of eight measures-related classifications. It is noted that there are many ways of classifying a program evaluation

effort: by subject matter of the evaluation; by the purpose of the evaluation; by the methodology employed in the evaluation; or by some other criteria.

Companion et al. (October 1977) discuss what they feel to be two ignored issues when developing a task taxonomy. The first is the set of criteria, i.e., rules on which a judgment can be based for the evaluation of how well a task taxonomy accomplishes the goals underlying its development. The second issue is the relationship between the taxonomic structure and empirical data (i.e., laboratory and field data). They list the following nine criteria which they feel should characterize a task taxonomy:

- It must simplify the description of tasks in the system.
- It should be generalized.
- It must be compatible with terms used by others.
- It must deal with all aspects of human performance in the system without logical error.
- It must be compatible with the theory or system to which it will be applied.
- It should help to predict operator performance as it is necessary to evaluate and compare performance between operators between different as well as identical tasks.
- It must have some practical or theoretical utility.
- It must be cost effective.
- It must provide a framework around which all relevant data can be integrated.

In still another extensively structured approach, Miller (1978) examines all biological and social systems and divides them into seven hierarchical levels: cells. organs, organisms, groups, organizations, societies or nations, and supranational systems. He identifies 19 critical subsystems and defines 13 distinct concepts which he feels must be understood in analyzing any living system at any level. The 13 concepts are: space and time, matter and energy, information, system, structure, process, type, level, echelon, suprasystem, subsystem and component, transmissions in concrete systems and, finally, steady state.

Siegel et al. (1977) developed a battlefield language taxonomy. Fifteen factors were identified which represented the perceptual substate of the Army field information linguistic system. The results of this study indicated that intelligence analaysts can classify messages reliably within the taxonomy. In addition, a computer system for the automatic classification of battlefield messages was presented.

O'Connor et al. (1977) presented an aircraft system inventory hierarchy which provided an hierarchical evaluation structure relating all the tests and evaluation information to the mission of the aircraft system under consideration.

Meyer et al. (1978, Vol. ~' developed a taxonomy of tactical flying skills in this study. It was developed as a user-oriented, skill-task analyses system for practical application in solving tactical Air Command continuation training problems and provided a behavioral data base for skill maintenance and reacquisition training research and development.

Cunningham (1978) described a basic systems model which is applied to evaluate organizational effectiveness and deals primarily with subsystem interrelationships. Basic to the model is an analysis of environmental inputs, methods by which the inputs are transformed (throughputs) and the end products of this transformation (outputs).

The philosophy underlying the study by Kaplan et al. (1978) is that understanding of missions is basic to the measurement of systems in operational tests. The authors assert that there must be a logical link between the missions to be performed and the selected measures of performance. The procedure followed in this study to accomplish this linkage is to define systems according to their generic class(es) and then define each generic class by general functional and hardware similarities. It is observed that systems belonging to the same generic class have certain missions in common while having other missions specific to themselves individually.

In another study, Uhlander (1970) described jobs by means of a taxonomy containing cognitive variance (responses more objectively characterized) and noncognitive variance (responses less objectively characterized). It was noted that the systems measurement bed assists the researcher in dealing with the different measurement characteristics of the two classes of jobs.

Cunningham et al. (1965) discuss the historical controversy concerning the measures used to assess performance, some of which purport to evaluate functional units of the system, others which deal with subsystems, and still others which attempt to assess the behavior of the total system. In the authors' view, little is known about the relationships among the various measurements or their relevance as criteria for making adequate judgments regarding operations. It has been asserted that single performance measures are inadequate for making overall evaluations of system effectiveness. However, the authors feel that combining measures into overall indices has, so far, not seemed to be of much help because the relationship between them is not often clearly understood. The authors feel that combining these measures does not necessarily improve the quality of system evaluation.

Finally, the need to provide operationally defined terms of behavior to compare the man-machine behavior of one system with another is noted by Meister et al. (1965). The authors suggest that a crucial characteristic of a system is its purposiveness (goal-directed behavior). It is stated that there are two groups of goals—mission oriented and supporting. Mission oriented goals seek to accomplish the system mission and direct the performance of all mission-related system activities. Supporting goals, on the other hand, seek to maintain the integrity of the system until the mission has been accomplished. The authors present in this a graph of the taxonomy system.

As might have been expected, little was found in this literature review which provides well-developed, practical techniques and convenient methods for

for system measurement. As can be noted from the above, work has been done in advancing some areas of taxonomy development. The major effort would still appear to have been made by Finley and her colleagues.

3. Overall Conceptual Process Model (CPM)

As can be seen in the previous section, the Systems Taxonomy Model deals with the contextual components of system measurement. The Overall Conceptual Process Model (CPM), however, is concerned with the entire systems measurement process. It is viewed as a systematic structure composed of four major subprocesses or components:

- Contextual Components (or STM): system, mission, and environment definition, constraints on the system, performance requirements of the system and performance criteria.

- Analytic Components: the attribute measures and the measurable attributes, the specific requirements, criteria and measurement procedures.

- Planning Components: analytic methods, parameter determination apparatus and personnel for testing and test plans.

- Application Components: test implementation, data analysis findings, and conclusions and recommendations.

The review of the literature provided a great deal of material with regard to this topical area. These reports contained much theoretical discussion of the subject as well as descriptions of the practical application of the measurement process. An attempt is made here to summarize as briefly as possible a representative sampling of the work which has been performed in this area.

One selected structure of the overall measurement process is provided by Simon (1974). The step-by-step procedure is summarized as follows:

- Review of documentation
- Formulation of test objectives
- Selection of applicable test concept
- Measures of effectiveness
- Test design
- Simulations
- Data
- Range instrumentation
- Test plan
- Conduct of test
- Data analysis

- Conclusions and recommendations
- Test report

Similarly, the Weapons System Effectiveness Industry Advisory Committee (1965, Vols. 1 and 2) determined that system evaluation can be reduced to the following ordered set of tasks:

- Mission definition
- System description
- Specification of Figures of Merit
- Identification of Accountable Factors
- Model construction
- Data acquisition
- Parameter estimation
- Model exercise

Meister (1978) identified server aspects of the measurement process that are common to any analysis. They include:

- Assessment of the impact of system parameters on personnel performance.

- Assessment of the impact of human factors on system outputs.

- Specification of the "mission scenario" of the system (initial stimulus to end-point).

- Replication (validation) of the research study under identical or simulated conditions.

It was further stated that all system-relevant factors must be included in any measurement situation with two factors involved: all variables affecting system output must be included, while all interactions must be included in the researcher's system representation (ensure that those variables chosen represent the operational system).

Bond et al. (1959) outlined a basic assessment/evaluation method as follows:

- A clear statement, in observable terms of the expected results of the treatment, including the time span over which a specific result can be measured.

- Development of relevant, reliable yardsticks (MOEs) which measure progress toward the stated objectives (expected results).

- Application of the yardsticks within the time spans of the objectives.

- Establishment of an evaluation design allowing the treatment effects to be distinguished from intervening contaminants.

- Establishment of the kinds and sources of information required to evaluate the treatment in terms of the objectives.

- Specification and examination of underlying personality and situational factors which explain the identified change.

Waag et al. (1975) stated that the implementation of a measurement system requires:

- Definition of criterion objectives in terms of a candidate set of simulated parameters.

- Evaluation of the proposed set of measures for the purpose of validation and simplification.

- Specification of criterion performance by requiring experienced instructor pilots to fly the particular maneuver.

- Collection of normative data using students as they progress through the training program.

In an investigation conducted by the U.S. Army Combat Developments Command (1968), the following procedures were undertaken:

- Creation of a data base
- Analysis of systems concepts
- Expansion of a data base
- Examination of suitable models
- Review of the personnel subsystem
- Critical incident analysis
- Measurement of man's contribution to system effectiveness

The process of the development of valid field performance measures, proposed by The Bunker-Ramo Corporation (1965), is as follows:

- Select tasks which manifest a range of behaviors from complex to simple which are related to total system function and which have a wide variety of operational conditions.

- Analyze tasks to describe: task hierarchy, their interrelationships, behavioral functions and the points at which work load conditions arise for personnel, etc.

- Develop and administer a task performance characteristics scaling test.

- Develop predictive criteria to be validated in field exercises.

- Conduct validation tests.

Five major evaluation phases were reported by McKendry et al. (1964):

- Preparation and initial planning
- Devising and writing the test plan
- Conducting the test
- Evaluation of data from the test
- Derivation of conclusions and recommendations of the final report

From the definition of a system to a quantitative criterion of its value, the following steps are identified by Harrision (1966):

- Define the mission as broadly as possible, being consistent with some concept of how its ability to achieve the former can be expressed quantitatively.

- The system designed to accomplish the mission should be explicitly defined to some "boundary." The latter must separate the system from its environment; contributions from other elements or systems are incidental.

- A criterion for judging the value of a system must be formulated and/or,

- A method of optimizing the design or choice of system devised.

- Based on the method of optimization chosen, certain types of measurements must be obtained for a complete set of characteristics at the highest level possible.

- A method of expressing the effectiveness of a system as a function of the elements in a set must be designed, or if this measure can't be obtained with the desired confidence of correctness, then

- The mission should be redefined such that effectiveness can be more confidently expressed.

Timson (1968) suggested the following steps in an evaluation of a total system's performance:

- Find design equations that relate the subsystem properties to the total system performance.
- Determine the subjective probabilities for the subsystem and the component properties that influence the total system performance.
- Utilize the Monte Carlo procedures to generate probability distributions for the system performance characteristics.
- Compute the statistical measures of the system performance probability distribution.
- Compare the statistical measures for the different time periods to obtain indications of progress.

The following steps, described by Duning et al. (1972), represent the blocks in a procedural block diagram culminating in system- and pilot-centered evaluation criteria:

- Describe vehicle operational profile.
- Select outcomes of interest.
- Specify outcomes and pilot acceptance in terms of critical limits of pertinent variables in numerical terms.
- Determine system error and state variable performance response to inputs.
- Determine outcome probabilities and pilot acceptance probabilities.
- Define safety, operational capability and pilot acceptance design qualities.
- Determine procedural variables.
- Determine task variables.
- Determine environmental variables.
- Define normal/degraded feedback arrangements and control-display mechanizations to perform functions. Allocate functions to manual and/or automatic systems.
- Identify system performance-centered variables and physical characteristics.
- Identify human operator-centered variables.

In the work reported by the McDonnell Douglas Astronautics Company, Eastern Division (September 1969, Book 2), it was stated that systems research has four stages or purposes with three kinds of research functions. The four stages are:

- Delineation of system requirements
- Delineation of design consequences, consequences of requirements
- System development and integration
- System evaluation

The three types of research functions are:

- Development of models
- Collection of research information
- Synthesis of information

The generic classification or indexing of the system is the first step in the human resources test and evaluation process (Kaplan et al. 1978). The subsequent steps are as follows:

- Assignment of missions
- Specification of system performance issues
- Identification of human performance functions and measures
- Identification of test conditions
- Specification of human resource issues and measures
- Operational testing
- Evaluation of operational testing results
- Diagnosis of performance inadequacies

Hutchins (1974) and Jahns et al. (1972) described the use of the Computer Aided Function Allocation and Evaluation System (CAFES). The program's principal objective is to facilitate application of essential elements of human factors technology in systems development using automatic data processing techniques to analyze and evaluate crew subsystem performance as it affects total systems effectiveness. Another computerized method is described by DiGialleonardo et al. (174). The technique for Interactive Systems Analysis (TISA) is a computerized technique for conducting systems analysis in a conversational model from interactive terminals.

Finally, several authors desdribed modelling techniques, including Hakanson (1967) who reported on an adaptive model of the development process in weapons systems which dtermines which tests, if any, should be performed at a given stage and if corrective actions should be taken. Topmiller (1968) discussed three research approaches to the problem of mathematically representing human performance parameters in various weapon systems, and Phatak (1973) studied the problem of developing realistic models for weapon system controllers that can be used to predict the effectiveness of manned weapon systems under stress conditions.

4. System Definition

Many of the articles, reports and books reviewed defined systems in terms of the particular hardware (e.g., tank or aircraft) that was at the focus of their study. Some defined systems from human perspectives (e.g., maintenance training or activities during a simulated task), although both the hardware and human perspectives also considered the procedures to accomplish the system activities. Still others, a few, discussed what system definition means in a theoretical sense; these will be presented first in this section.

Miller, J.G. (1978) defines a system as a set of interacting units with relationships among them. The word "set" imples that the units have some common properties. These common properties are essential if the units are to interact or have relationships. The state of each unit is constrained by, or dependent on, the state of other units.

According to Cunningham (1978), characteristics of the system-model are physical and chemical laws that are applicable to social organizations in six ways:

- Every system uses energy in a cyclical way: the environment product or output becomes the energy source for the subsequent activity cycle.
- Systems are separated from their environments by boundaries; since events are structured in a systematic way in an organization, the boundaries of the system are between events.
- Equifinality in open systems: a final or specific end state can be reached by a diversity of inputs and varying enviromental and internal activities.
- Entrophy: in nature, all organized systems "wind down" or move toward disorganization and/or death—this is the second law of thermodynamics; in open systems, however, negative entrophy allows the system to temporarily circumvent enthrophy by importing more environmental energy than it expends.
- Equilibrium, or dynamic homeostasis: systems adapt to change and attempt to maintain a balance in their status quo; the system will also attempt to acquire a margin of safety in inputs above and beyond what it needs for mere survival.
- Feedback: an information input into the system resulting from previous outputs and their effect on the system's environment.

Churchman (1978) defined "system" as that which a decision maker can control and change, and U.S. Army Combat Development Command (1968) defined it as a conceptual framework for attacking problems. In its broadest terms, a system is comprised of hardware, facilities, logistic support, and the trained manpower required for operation in a particular environment. In a similar definition, Smith et al. (1969) (Vol. I) felt a system was a set of personnel-equipment functional units whose collective purpose is to achieve a particular goal. The life cycle of a system can be divided into the conceptual, acquisition and operational phases (Weapon System Effectiveness Industry Advisory Committee, 1965, Vol. 1).

A multi-model system is described by Bond et al. (1970) as a collection of tasks functionally connected by independent subsystems. Each subsystem is a set of identical functional groups of a given type. TAC Fire for example, is a multi-purpose system but it also multi-functional (Kriske et al. 1977). Its team members participate in all functions with the same or similar team dimensions.

According to Smith et al. (1969) (Vol. 1), differentiating between systems and subsystems is arbitrarily based since most systems can be defined as subsystems when referenced to larger overall systems of which they are a part. What is important is the relationship of a given system's goals with respect to those of another. Thus, the effectiveness of a given system should be evaluated with respect to the parent system. A system can be assumed to have been 100% effective if it performed up to its maximum capability, regardless of whether or not it was subsequently destroyed. Thus, capability is an important factor contributing to establishing system effectiveness requirements and the evaluation of system effectiveness.

Specification of system states tends less to simply imply transitory methods for achieving those states and tends to lead to a creative, open-minded approach to analysis, both for new designs and for evaluation of existing systems (Mitchell et al. June 1967). An understanding of systems is necessary for both the evaluation of a particular system and for the development of systems evaluation methodology. Further, review of documentation relevant to a given system to be tested is important because of the test manager's need to have a clear understanding of the critical issues, data requirements and test objectives, and need to be familiar with the system, how it operates and its history in previous tests (Simon et al. 1974).

In this review, several of the systems described were measurement systems. For example, the CAFES was developed to provide an integrated system of computer models which progress from the early concept formulation phase through crew station design to the final test and evaluation of the completed product (Hutchins, 1974). Another example is provided by Dunlap et al. (1967) who developed requirements for an instrumentation system to measure automatically the performance of test participants. Requirements and specifications were developed for a centralized, computerized, data logging procedure to record, process and statistically anlyze performance data collected by the system. On the other hand, the SEA System (Polak et al. 1974) has two major functions: 1) to check out, control, monitor and perform statistical analyses associated with tracking simulators; and 2) to provide an estimation of weapon system effectiveness. Two other measurement systems have been utilized for specific evaluation. One, according to the Labor/Management Task Force on Rail Transportation (1975), was a performance measurement system for the St. Louis (railroad) Terminal Project. The other system's purpose (Rasch, 1973) was for making tradeoffs to specify those performance measurement factors relevant to ship acquisition and to specify standards for using TPM outputs to make the necessary tradeoffs for ship design decisions.

A number of studies concerned with defining systems dealt with human task performance with specific equipment. For example, two studies (Klein et al. 1969 and Klein, R. undated) defined their systems as being composed of infantrymen, their weapons and equipment. Another system was concerned with rifle squads, rifle platoons and tank platoons (Clovis et al. 1975). Still another combat "system" force was a cavalry unit (U.S. Department of the Army, 1977). A prototype handbook, as described by Kaplan et al. 1978), lists 11 generic class of Army systems.

The Defense Satellite Communications System was described (Ray et al. 1979) as was the AAW Combat Information Center (Smith et. al 1969, Vol. 2), Airborne Warning and Control System (Turner et al. 1972), the AN/TSQ-73 Missile Minder System and the SAINT System (Wortman et al. 1979), the SAGE System (Mitchell et al. June 1967), and others (Weapon System Effectiveness Industry Advisory Committee, 1965, Vol. 3; Wellman et al. 1972; Beau, 1964; Hicks, October 1977).

Several of the systems that were defined were concerned with aircraft and pilot performance (Rhoads, 1970; Meyer et al. 1978, Vol. 1; Topmiller, May 1968; Matheny et al. 1971; Kiraly et al. 1970) as well as pilot training and simulation (Waag et al. 1975; Grunzke, 1978; Campbell et al. 1977; Connelly et al. December 1974, AFHRL-TR-74-88; Vreuls et al. 1975; Hill et al. 1974; and Irish et al. 1977).

A number of other systems that were described in articles were training systems. These educational systems ranged from those used in training air traffic controllers (Buckley et al. 1976), to inventory and materiel facilities training (Hansen et al. 1977), to sonar operations (Fischl et al. 1968), to operator loading in man-machine systems (Siegal et al. undated), to a closed loop system (Akashi et al. undated), and others (U.S. Army Infantry School, 1976; Mitchell et al. 1967; Hall, 1973; Goldbeck, 1971; Mumford et al. 1961; Gustafson, 1967; Willis, 1967; Siegal et al. 1961; Performance Measurement Associates, Inc., 1978; Boycan, 1972; Finley et al. 1976; Foley, 1975; Anderson, 1977; Siegel, 1970; Spencer, 1967; Thurmond, undated).

5. Mission Definition

A mission describes the man-machine activities performed to accomplish the primary systems goals (Meister, 1965). unless it is framed in terms of mission goals, system behavior becomes extremely difficult to explain or understand because purpose is the single factor which unifies a great variety of disparate system behavior. A system's required overall capability is directly related to its set of defined mission objectives (Chop, 1972). Further, the mission of the human component in a system is that his/her function is performed adequately and in such a way that it will lead toward mission accomplishment (Willis, 1967).

According to Mitchell et al. (June 1967), it is necessary that valid system effectiveness requirements exist and are derived from mission analyses, and that the system is partitioned into manageable units for evaluation of their contribution to system performance. Knowledge of the missions allows the analyst to specify the system performance issues of interest (Kaplan, 1978).

The Labor/Management Task Force on Rail Transportation (1975) said the mission of current performance measurements were found to be designed to support one or more of the following functions:

- Evaluate performance and trigger the planning process to develop changes that will produce improved performance.

- Evaluate experimental changes in operations to determine the actual improvement in performance.

- Monitor the operations to provide information that results in corrective action to prevent a deterioration in performance.

- Assess the performance of the managers responsible for the operations.

In applying a system model to real-world modeling of organizational effectiveness, Cunningham (1978) felt the mission of these organizations was:

- The organization's ability to respond to its external environment

- The organization's ability to utilize resources in producing outputs and maintenance/restoration of the system

- The organization's ability to bargain and optimize its use of resources in an environment with multiple decision-makers, each with different goals

Many of the articles, books and reports reviewed defined their particular mission in terms of mission objectives. Some of these defined their mission in very broad terms applicable across many specific systems. Others defined their mission in terms of human performance associated with particular hardware. Still others described the mission as strictly human factors (or psychological) processes.

For example, according to Geer (1977, D194-10006-2), *human factors* engineering, test and evaluation existed to:

- Demonstrate system conformance, equipment and facility design to human engineering design criteria.

- Determine man's contribution to performance requirements.

- Quantify man-machine interactive measures of system.

- Detect undesirable design on procedural feastures of system.

Another study (von Winterfeldt, 1975) defined the mission of utility theory as its application to certain decision situations that may be classified according to three factors: static versus dynamic decision environment; single decision makers versus multiple decision makers; and single aspect choice entity versus multiple aspect choice entity. Another human factors type mission definition was the CAFES objective (Hutchins, 1974). This allows the human factors engineer to treat in a comprehensive way all parameters to be considered in the designing of a man-machine interface of advanced Navy systems. The purpose of still another study (McCalpin et al. 1974) was the development of the general procedures necessary to obtain human performance data which will satisfy a prior model that includes human performance data in models of infantry weapon system reliability. Other studies (Brokenburr, 1978; Breaux, 1976; Connelly et al. 1977; and Meister, 1978; for example) defined training and problem solving missions.

There were very general mission definitions too. The mission of the system was the most effective use of men, equipment and weapons in a combat situation, for example (Klein et al. 1969). According to Swink et al. (1978), the mission definition was the effective operation of the aircraft on a typical mission. The maintenance man's mission in a man-machine system is to ensure that the machine subsystem is in prime operating condition when the mission is started (Foley, 1975). Pritsker et al. (1974) reported that a mission was defined as a network of tasks performed by a crew of operators having a complement of equipment in the face of environmental factors. Other rather general mission definitions were specified by Rasch (1973), Williams et al. (July 1975) and the U.S. Department of the Army (1967).

A number of studies that were reviewed involved military systems. For example, the mission described by one report was the use of small arms in combat situations by infantrymen (Klein, undated). The missions, according to Clovis (1975), are selected kinds of engagement with the enemy for each system. This mission of this system was to score a first-round target hit in the minimum possible time (U.S. Department of the Army, March 1977). The purpose of the recoilless weapon system was to provide a comprehensive summary of the available relevant technology and the system engineering rationale (U.S. Army Army Materiel Command, 1976). The cavalry's basic mission, as stated by the U.S. Department of the Army (July 1977), is reconnaissance and security. The mission of each howitzer section was uncoupling the howitzer, preparing for action, firing and march order (Dunlap and Associates, Inc., 1966). Other military missions are described in Ultrasystems, Inc., 1972, Vol. 1; Wellman et al. 1972; Chasteen et al. 1975; Andrews, 1977; Jaschen, 1975; and Dunlap et al. 1967).

Communication can be a mission. For example, the primary missions of the DSCS was an increased communications capability, particularly an improved ability to operate in an electronic warfare environment (Ray et al. 1979). In another system, the CE-75 system, the mission was to provide a means for the timely transfer of meaningful and significant information from action officer to action officer (U.S. Army Combat Developments Command, 1968). Another report (Weinstock et al. 1969) described the purpose of the systems to transfer information between two separate locations.

The remaining several studies described in this section of this report define the mission of aircraft systems, aircraft activity and aircraft performance training. For example, the mission of the system was to perform high and low altitude all-weather attackes (Campbell, 1977); the aircraft mission was the engagement in precision weapon delivery or air-to-air combat (Phatak, 1973); the mission to be evaluated was aircraft approach and landing using MLS (Duning et al. 1972); the purpose of the vehicles was the resupply of an orbiting space station in a 300 nautical mile orbit (McDonnell Douglas Astronautics Company, 1969, Book 1); the mission of this system was to provide aircrews with a safe, reliable and compact oxygen system (Kiraly et al. 1970); the mission was the evaluation of five pilot training maneuvers (Connelly et al. December 1974, AFHRL-TR-74-88); and others (Hyatt et al. 1975; Vreuls et al. 1975; Irish et al. 1977; Rhoads, 1970; and Turner et al. 1972).

6. <u>Environment Definition</u>

During a system evaluation, the environment under which the test takes place will have, in most cases, considerable impact on the validity of the results

of the study. Churchman (1971) views the environment of a system as a set of things which the decision maker cannot control but which, nevertheless, affect the performance of the system. Levy (1968) notes the need for collecting performance data in field situations to validate laboratory studies. Meister (1968) asserts the behavioral models characteristically employ laboratory data and have ignored or have been unable to handle natural event data. In the research reviewed, many researchers recognized the difficulty in achieving real-world conditions in a simulated situation.

Other researchers noted that there were problems in simulating real-life conditions. Klein (in preparation) said that the inability to duplicate combat actions and tasks in a test facility affected the validity of test results in his study. U.S. Army Operational Test and Evaluation Agency (1976) stated that time constraints did not permit waiting for the desired winter conditions and that as the intent of the study was specifically to address wetness effect on system functions under winter thaw conditions, this constraint presented a problem. Some researchers made considerable efforts to conduct their tests under realistic circumstances. In the Dunlap et al. (1967) study, the experiment took place at an integrated test facility consisting of eight test situations which had previously been developed and tested. In an aircraft evaluation, three levels of wind, three levels of turbulance and two levels of ceiling visibility were simulated (Irish et al. 1977). Thermal stress conditions were simulated in studies conducted by Repperger et al. (1978) and both the U.S. Army test and Evaluation Command (1970) and the U.S. Army Infantry Board (1971) attempted to duplicate the physical and environmental conditions to be found in the equipment's future use. Rhoads (1970) simulated different conditions of static and dynamic characteristics in the B-1 type airplane and Spyker et al. (1971) reported that although the test took place in the laboratory situation, the physical, psychological and environmental conditions were kept as constant as possible. Brown (1977) felt that the tests associated with his study were conducted with as much tactical realism as possible and included operation on primary, secondary and cross-country terrain. Featherstone et al. (1975) conducted tests on a typical pistol range common to most Army installations, and the subjects participating in Dunlap and Associates, Inc. (1966) study were provided with full tactical uniform and live ammunition and emplaced on an actual field site on the firing range.

In summary, in much of the research reviewed, efforts were made to simulate "real-life" conditions in the test situations. However, many researchers experienced difficulties in this area and few provided detailed descriptions or definitions of the environments, whether real or simulated.

7. General Constraints

As stated previously, the analyst needs to know all of the limitations and conditions that will be imposed on the operating system (as opposed to the evaluation effort) so that realistic and appropriate measurement can occur.

In some of the studies reviewed, the researchers reported on such system limitations. In the study conducted by Hyatt et al. (1975) on the microwave landing system, it was noted that aircraft using the same airspace are

likely to have a wide spectrum of data processing capabilities. Different aircraft have different limitations regarding equipment factors such as weight, space, capital and maintenance costs.

In a different kind of example, Gustafson (1967) describes a personnel system which has limitations caused by operator retention problems, which may have implications for the measurement process. The Air Force competes with industry for the services of trained personnel and is not usually able to retain them for longer than a 4-year tour of duty. Therefore there is a need to train maintenance personnel at a great variety of tasks, some of which require high levels of skill, in a short period of time without further education, cross-training or retraining.

The U.S. Army Combat Developments Command (1968) reported that a man-machine interface (MMI) is a boundary at which a man and a machine interface in order to acheive a system objective. The extent of this boundary is constrained by three factors: 1) tasks required of both the man and the machine to attain the system objectives; 2) capabilities and limitations of the machine; and 3) system objectives as affected by environment, personnel policies and equipment use.

Karaush (June 1969) noted problems in planning and estimating the system's workload due to a variation in demand and the random user times. Meister (1967) perceived problems in predicting performance. He said that it is necessary to account for the fact that more than one task may be performed concurrently by the same operator. Each of the two concurrent tasks has its own important parameters for predicting the task's individual performance, therefore this concurrency is another important parameter for measuring and predicting total performance.

Generally, however, in the work reviewed, it appears that there is little or no discussion of the limitations of the system itself and the implications on the measurement process, although much is reported on the limitations of the evaluation study.

8. <u>Performance Requirements, Ultimate</u>

In most cases, the ultimate performance requirement of any system is that it performs its mission. Chop (1972) states that system capability is a focal parameter in that it is the top performance parameter of a system against which all other parameters are funneled, evaluated, cross traded and optimized. It provides the link up of system performance with mission objectives. Mitchell et al. (June 1967) states that to define an acceptable level of performance with respect to system objectives, a stipulated value is established on the performance dimensions, and that value constitutes the System Engineering Requirement (SER). Effectiveness requirements may take the form of a single value on an effectiveness dimension, or several values, or an interval may represent levels of effectiveness which are acceptable under specified operating or environmental conditions. When more than one effectiveness dimension is needed to reflect the system objective adequately, the SER may be represented as an index resulting from the mathematical combination of values on several effectiveness dimensions. For allocation of SERs,

mission analyses must have been directed toward defining requirements appropriate for effectiveness analyses. Values along all relevant dimensions must emerge as an end product. Currently such end products are sorely lacking due to the intuitive approach to design for meeting imprecisely defined system objectives. SERs are rarely specified, either because 1) they had not been considered, or 2) the researchers don't wish to face the fact that serious objectives may not always be reached, or 3) are unwilling to record fallibility.

Anderson (1977) states that an aircraft system often has more than one requirement and in his study he says that the utility of the system depends upon kill potential, probability of reaching the target, probability of survival and availability. He feels that the aircraft's worth cannot be assessed by considering these functions in isolation.

Often the mission can be simply stated. For example, in the evaluation of a small arms weapons system, the objective is to close with and defeat the enemy (Klein, 1969) or in the case of a military unit, the objective might be to destroy the enemy's ability to wage war (Clovis et al. 1975). When the evaluation is concerned with a specific segment of an overall military system, then the ultimate performance requirement, as would be expected, is narrowed down to the segment of interest and might be expressed as the requirement that aircrews receive the life support necessary during flight (Kiraly et al. 1970), that the system provide accurate information regarding an aircraft's position during a landing approach (Hyatt et al. 1975) or the safe and expeditious movement of aircraft through a sector (Buckley et al. 1976).

In a study by Weinstock et al. (1969), the ultimate requirement is that the information is received and reaches it destination and Siegel et al. (1961) describe the mission of the training program as the preparation of students for the jobs involved after training.

It appears in the reports reviewed that few authors actually describe the mission of the system which is being evaluated in any detail. The above represents an effort to describe briefly some of the ultimate performance requirements with which these evaluations were concerned.

9. Performance Criteria, Ultimate

The ultimate performance criteria, as defined for use in this study, is the criteria, or standards upon which one can measure whether or not the system performs its mission. Topmiller (1968) defined the major parameters of system effectiveness as availability, capability and dependability. Availability is equivalent to the system's readiness to perform its mission; capability is the measure of a system's ability to achieve its mission objectives, and dependability is the measure of the system's condition at points during the mission; and that these parameters are criteria of system performance which require measurement and prediction. Chapanis (1970) states that the value or worth of a system is normally judged by several criteria not necessarily all compatible. Typical man-machine system criteria include: 1) anticipated system lifetime; 2) appearance; 3) comfort; 4) convenience; 5) ease of operation; 6) familiarity; 7) initial cost; 8) maintainability; 9) manpower requirements; 10) operating cost; 11) reliability; 12) safety; and 13) training requirements. Bond et al. (1970) report that there are only a relatively few indices that can be used as criteria for evaluating learning. They are: 1) high

degree of accuracy in performing the learned response; 2) significantly shorter reaction latency than at the beginning of practice; 3) increased rate or speed of correct response; 4) increased amplitude of response; 5) increased resistance to experimental extinction; 6) increased resistance to retroactive inhibition from subsequent learning compared to the amount occurring when learning stops short of mastery; 7) increased positive transfer to subsequent learning in similar situations; and 8) a degree of generalization to similar status events.

In training programs, one measure of success would be the performance of the trainee at the end of training (Obermayer et al. 1974, Phase I). The ultimate criteria for the MLS system is that the aircraft must be within a successful landing window as defined by dispersions at decision height and reference position at touchdown (Duning et al. 1972). Munitions effectiveness for a single round is defined by Williams et al. (July 1975) as effectiveness equals availability, reliability and effective coverage. If the objective of a system is safety, one measure is the number of accidents which occur (Henderson et al. 1973) and in an information system, the criteria is that the information be understood within acceptable boundaries of quality and error rate, and that it reaches its destination in a timely fashion (Weinstock et al. 1969). In combat situations, the criteria was the achievement of a hit during a quick-fire engagement in the shortest period of time (Klein et al. 1969) and the number of enemy casualties was one of the criteria for success in Klein's (undated) study. Finally, the criteria used by Siegel et al. (1961) was that the systems and equipment be maintained in a state of readiness and that the mission be completed in a minimum time with appropriate levels of accuracy and reliability.

10. Practical Measurable Attributes

Researchers provided an abundance of information on the practical measurable attributes in their studies. However, it was more difficult to determine what their rationale was during the attribute selection process. Miller (1978) covers the subject of system measurement and variable selection and provides critical comment on some of the work in the field. Mitchell et al. (1967) state that quantification of effectiveness requires the identification of one or more measurement dimensions. Most frequently used measurement dimensions are accuracy, time quantity and rate constrained by cost limiations; and effectiveness dimensions must be related as directly as possible to the stated system objectives.

Irish et al. (1977) reports that because skillful piloting involves the attempt to maintain or change to specified flight parameters, deviations from these desired parameters provide quantitative objective performance measurements. Typically, in the literature reviewed, flight parameters such as altitude, air speed, headings, pitch and rol rate, range, etc., were the measured attributes (Waag et al. 1975 and Timson, 1968). In an evaluation of a pilot's performance during a microwave landing approach, Hyatt et al. (1975) also used deviations in position and speed from the planned glide path as an effectiveness measure.

In some of the studies reviewed, researchers depended entirely upon data which can be measured either quantitatively or qualitatively. In other work, researches utilized both types of performance measures in the evaluation. For example, Duning et al. (1972), in their research on control-display testing requirements, described evaluation criteria which were commensurate with absolute values such as location with respect to approach window, location with respect to runway

at touchdown, etc. Qualitative assessments were made with regard to areas of subjective evaluation such as missed approach procedures and failure detection procedures. Rhoads (1970), in his evaluation of four cockpit controller configurations, used qualitative data by obtaining pilots' inflight comments and ratings of handling characteristics and tracking error. Fineberg et al. (undated) determined mission success by analyzing the instructor's subjective rating of the trainees' navigational success and by constructing an objective measure in terms of the number of landing zones missed, etc.

Burington (1961) suggested that an analysis of the potential effectiveness of a typical weapon system is concerned with such facts as the ability to detect, locate and identify, designate and track the target. Also the ability to bring the weapon into range and place the missile within the desired damaging radius of the target and the ability to detonate the warhead at the proper place, manner and time were all appropriate measures of the success of a mission. Other factors could include the ability to inflict the quality of damage desired, rapid repeat fire and the number of targets that may be engaged simultaneously within a given interval of time.

The criteria used to evaluate the performance of infantrymen using small arms weapons (Klein, 1969) were grouped into four areas for purposes of this evaluation: accuracy, sustainability, responsiveness and reliability. Klein et al. (1969), in another small arms evaluation, prepared a list of 26 separate combat actions and a list of tasks normally accomplished by the infantryman when executing combat action. It was determined that three basic tactical situations (attack, quick-fire and defense) would accommodate all of these actions and tasks. Chapanis (1970) listed some common ergonomic and human factors research dependent measures used to assess system performance. They included:

- Accuracy
- Cardiovascular response
- Critical flicker fusion
- EEG
- Energy expenditure
- Muscle tension
- Psychophysical thresholds
- Ratings (of comfort, annoyance, etc.)
- Reaction time
- Respiratory responses
- Spare mental capacity
- Speed
- Trials to learn

The Weapon System Effectiveness Industry Advisory Committee (1965, Vols. 1 and 3) determined that a system's effectiveness can be measured in terms of its availability, dependability and capability. Availability is defined as the system condition at the start of the mission and is a function of its relationship between hardware, personnel and procedures. Dependability is a measure of the system conditions at one or more points during the mission, and capability is a measure of the ability of the system to acheive the mission objectives. Capability therefore accounts for the performance spectrum of a system.

The above represents a sampling of the types of data collected for measurement in the research reviewed. Obviously the spectrum of the type of measurable attributes is larger than reported here and, as stated earlier, generally there is little information or justification of why the particular variables were selected for measurement.

11. Practical Attribute Measures

Once the measurable attributes have been selected, the next step in the process is to determine how these attributes will be measured. Various statistical and mathematical techniques were utilized in the research reviewed to obtain both quantitative and qualitative assessments. A description follows of the scaling methods used in the measurement process by many of the researchers.

Observations of unordered variables are one of the most primitive forms of measurement and are described as constituting a nominal scale. An example of a variable in which the observations constitute a nominal scale would be individuals classified by sex. Only two values of this variable are possible, a male and a female, and the basic data would thus consist of the number of observations in each of the two classes—male and female. The data resulting from nominal scales are often referred to as categorical data, frequency data, attribute data or enumeration data.

Ordinal, interval and ratio scales are all relative measures. In ordinal scaling, the observations may be ordered in such a way that one observation represents more of a given variable than another observation. By comparing the height of, say, five individuals and assigning the number five to the tallest, four to the next, three to the next, two to the next and one to the shortest, this observation would be described as constituting an ordinal scale, and the numbers used are called ranks.

When numbers are used to identify observations and not only represent an ordering of the observations but also convey meaningful information, with regard to distance or degree of difference between all observations, the observations are said to constitute an interval scale. Thus, if the numbers, 7, 5 and 2 identify three different observations, it tells us that 7 is 2 units greater than 5, and that 5 is 3 units greater than 2, etc.

A ratio scale is an interval scale with an absolute zero. Length as measured in units of inches or feet is a ratio scale, for the origin of this scale is an absolute zero corresponding to no length at all.

Nominal and relative scales of measurement are related to measurement systems by Finley et al. (1975). They describe a Systems Taxonomy Model consisting of three major levels: 1) system objectives (the reasons for a particular systems existence; 2) system functional purpose (that which it must achieve to some level of adequacy); and 3) system characteristics—structural, operator/equipment, operating and support requirements (how the system is to or does operate). The definition of these three model levels includes a relationship to the nominal vs. relative levels of measurement. Typically, nominal system measures are related to the system objectives, nominal and relative measures are related to the system functional purpose, and relative system measures are related to the system characteristics.

The authors state that in the interest of performing studies which will gradually form a systems and system component relationships information base useful to analysts and practitioners in solving applied systems problems, it is recommended that the researchers always start at levels one or two and be sure to include all of the lower levels.

Miller (1969) described performance measures as what an alternative can deliver and the performance criteria as what the decision maker desires. He said that performance measures should be selected for each of the lowest level criterion and that the purpose of selecting performance measures is to establish concrete connections between desires and deliverable performance from real alternatives.

A reference source of measures of effectiveness used in Naval warfare and previous projects is presented in Rau's (1974) handbook. In a discussion of the choice of MOEs to be used in an evaluation of a system at any level, the following basic requirements are listed by the author. When selecting an MOE:

- It must directly relate to how well the specific objective is met.
- It should be relevant to the mission or operational role of interest.
- It should be precisely defined and expressed in terms meaningul to the decision maker in order to prevent decision makers and others from misunderstanding the implications of the MOE.
- It must be capable of exact quantitative definition in terms of inputs that are measurable. If the inputs are not measurable, the MOE cannot be evaluated.
- It must be feasible to measure or calculate.
- It should have exhaustive inputs and be sensitive to all variables and factors affecting the item (i.e., platform, system, subsystem or equipment). By this it is meant that anything that affects the item's effectiveness should appear as an input to the MOE in some fashion. This assures that all aspects that can affect the item's effectiveness are included in the inputs.
- It, as well as its inputs, should be mutually exclusive in the sense that no aspect should be "counted" more than once.

An appendix to this handbook provides a measures of effectiveness data base derived from OT&E Projects. For each Naval system or subsystem discussed, there is a description of the system, the specific objective(s) of the evaluation and the appropriate measures of effectiveness. For example, for a UHF Transceiver which is part of a communications system, the objective is to determine the adequacy of voice communications for both plain voice and secure voice. The measures recommended are: 1) mean error rate which is defined as the number of words missed

per 25-word message; 2) the probability that a rhyme word transmitted by this system in correctly interpreted; and 3) the percent sentence intelligibility.

Miller (1978) discusses in his book, system and subsystem indicators. He states that many subsystem and systemwide variables fluctuate constantly in every living system, and that if the changing values of conceptual variables are to be measured in a concrete system in space time, an observer or scientist must use some measuring instrument or technique—that is, an indicator—to do so. There are many kinds of such indicators and those used in studies at one level of system may be different from those used at another level. The author discusses system and subsystem indicators and suggests that various sets of organizational indicators have been devised. Some of the indicators are precisely quantifiable, others depend upon more subjective evaluations, like responses to questionnaires. A list of organization indicators include the following:

- Personnel indicators (number of people, types, ages, types of training, etc.)
- Product or service indicators (total output or processing capacity per unit of time, production time per unit, overhead cost per unit, customer satisfaction, etc.)
- Financial indicators (which are concerned with monetary information flows)
- Other indicators (such as lag between demand for services and response or amount of information processed per unit of time, etc.)

Hunter's (1976) method relied largely on a personnel reliability index modeled after an equipment reliability index. Specifically, the index was based on the compounding of probability of successful performance values for each of eight factorially derived job dimensions. A second instrument, based on Guttman-scaled checklist, was also described, and yielded an absolute measure of performance.

The following are the components and measures for which 19 scales were formed in a reliability study conducted by Farina et al. (1971):

Component	Measure
Goal	Number of output units
	Duration for which an output unit is maintained
	Number of elements per output unit
	Workload imposed by task goal
	Difficulty of goal attainment
Response	Precision
	Rate
	Simultaneity of responses
	Amount of muscular effort involved

Component	Measure
Procedures	Number of steps
	Dependency among procedural steps
	Adherence to procedures
	Procedural complexity
Stimulus	Variability
	Duration
	Regularity of stimulus occurrence
Stimulus-responses	Degree of operator control
	Reaction time/feedback lag relationship
	Decision making

Turner et al. (1972) measured reaction in units of time and surveillance ability in terms of the numbers of friendly/hostile aircraft detected, identified and tracked. Command was measured by the ability to allocate resources in terms of number and percent of sorties scrambled, immediate response requests accommodated and sorties diverted. Control reflects the number and percent of friendly aircraft under control per unit of time per sortie.

In their study of the SAGE system, Sheldon et al. (1967) translated the overall objective into three quantifiable criterion measures: 1) percentage fakers killed, 2) faker life time in system's air space, and 3) depth of penetration. These measures were supplemented by other measures concerning explicit system functions including detection latency, interception time and tactical action latency, etc.

A sampling of other attribute measures found in the literature reviewed and grouped arbitrarily is given below.

Aircraft Systems (Buckley et al. 1976; Dunlap and Associates, Inc., 1966; and Hyatt et al. 1975)

- Errors across track, along track, above and below glide path and speed along path
- Number of conflictions
- Number of delays
- Cumulative delay time
- Number of completed flights
- Cumulative air/ground communication time
- Number of aircraft handled
- Number of identifications required
- Number of aircraft in sample
- Number of completable flights
- Number of conflictions/number of aircraft handled

Aircraft Systems (Continued)

- Number of conflictions/number of delays
- Number of delays/number of aircraft in sample
- Cumulative delay time/number of aircraft in sample
- Number of completed flights/number of completable flights
- Number of contacts/number of aircraft handled
- Communication time/number of contacts
- Number of aircraft handled/number of aircraft in sample
- Correlation hold-delay transformation
- Number of identifications requested minus number of aircraft in sample
- Controller heart rate
- Number of scheduled and unscheduled oral communications
- Number of scheduled and unscheduled visual communications
- Number of errors by team and individual
- Number of unsafe conditions
- Time data in minutes and seconds
- Quality of data ("good," "fair," or "poor," based on the subjective judgment of an experienced battery officer)
- Heart rate of certain team members

Weapon Systems (Burington, 1961; Egbert et al. 1973; Klein et al. 1969; Klein, undated; Taylor et al. 1977; Ultrasystems, Inc., 1972, Vol. 2; and U.S. Army, Army Materiel Command, 1976)

- Probability of submarine detection by helicopter
- Time required for removal and installation of the firing port weapon
- Time to first round, time between trigger pulls, distribution of near misses, time to shift fire and hits per pound expressed as a percent of a soldier's basic load
- Probability that one burst of a missile will inflict "kill"
- Number of hits
- Probability of a kill given a hit
- Probability of a hit or hits on a target occurring out of a given number of rounds fired at a target
- Probability of submarine detection, localization and kills
- Ratio of the incremental improvement in accomplishing the mission to the incremental monetary cost of such an improvement
- Detection range of raid relative to the vital area center for a given intercept range

Weapon Systems (Continued)

- Expected number of targets destroyed in a given period of time
- Difference in fuel consumption due to the bathythermograph maneuver
- Total force level required to clear a given area in a given time
- Expected number of ships hit
- Elapsed time to target detection
- Maximum exposure time of the submarine

Training and Performance (Akashi et al. undated; Bond et al. 1970; Featherstone et al. 1975; and Ford et al. 1974)

- The time it took for an individual to accomplish these tasks without error
- Total time during which the error signal exceeded an arbitrarily chosen threshold
- Time to perform a task
- Gain scores (difference between post-test and pre-test scores)
- Process scores (assessment based upon application of procedures rather than overall success in problem-solving)
- Time to criterion (time required to complete some work or achieve some level of success)
- Error rate
- Persistence measures (staying with some specific training sequence)
- Transfer measures (generalizability of the learning to other situations)
- Time vs. achievement measures

Miscellaneous (Anderson, 1977; Beau, 1964; Harry, 1975; Matheny et al. 1971; and Ray et al. 1979)

- Labor cost, material cost, overhead cost, cost of waste, breakage
- Number of defects in finished product
- Bit error rate (BER), test tone-to-noise spectral density ratio (These measurements are used to determine if a communication link will pass data traffic.)
- Average absolute deviation from standard percent of time out of design limits (continuous control tasks)
- Percent of incorrect responses (discrete control tasks)
- Percent settings not on design setting or percent outside design limits (pointer/symbol positioning tasks)

Miscellaneous (Continued)

- Percent of decisions agreeing with judges' established decisions (technical decision tasks)
- Clearance rate of arrests through court system
- Percentage of individuals, by age groups, using a city recreation facility, circulation per capita of library materials
- Number and rate of injuries and deaths due to fire
- Subjective measurements of passenger comfort on local transportation
- Response time to handle complaints

12. Performance Requirements, Specific

This topical area is intended to describe the system's outputs, products or end results in terms specifically keyed to the selected measures. For example, the requirement of Kiraly et al. (1970) was that the system meet safety and comfort limits, and Siegel et al. (1974) deemed a training program effective if the graduate carries out the duties of his job proficiently. In Self's (1972) study, it was stated that an ideal observer would detect all of the targets, accurately distinguish between non-targets and targets, and detect and recognize the instant an image appears on a display. Similarly, Meyer et al. (1978, Vol. 1) described the specific application of the cues categories in performing a surface task analysis. Generally, the surface analysis must identify the aircraft type, maneuver the weapons delivery, determine whether the maneuver environment is a range or a tactically oriented one, determine flight paths, etc.

In a study to assess the adequcy of a number of organization's resources for coping with disasters of various magnitudes (Cunningham, 1978), the specific requirements of one of the factors (responding to the external environment) was to measure the organization's ability to achieve the highest resource allocation for various levels of damage. In a study of measures of effectiveness used in Naval analysis studies, a representative list of specific performance requirements included: detection of a submarine, successful attack capability, survival of aircraft and planting of mines, clearance of mine fields, surveillance and tracking of ships at sea, and time to prepare for attack (Ultrasystems, Inc., 1972, Vol. 4).

Transportability, mobility, capacity, quality, serviceability, and vulnerability were the specific requirements of the study by Weinstock et al. (1969). The Weapon System Effectiveness Industry Advisory Committee (1965, Vol. 2) specified the range at which a target should be detected and tracked within an admissible error rate. Cunningham's (1978) primary objective was to determine whether a system will operate under operational live user conditions while meeting requirements for reliability and response time.

Finally, verification of a radar's ability to meet stated operational requirements in terms of the radar's military ability, operational effectiveness and suitability were the testing objectives of the study conducted by Andrews (1976).

13. Performance Criteria, Specific

The specific performance criteria are the standards by which assessments can be made of whether a system meets its specific performance requirements. Miller (1969) states that having established a list of overall objectives, the second step is to generate a hierarchical structure of successively more specific performance criteria. This involves breaking down or subdividing higher level criteria into one or more lower level criteria

Holshouser (1977) states that to quantify technical performance, the following four steps must be taken:

- Identify performance variables (inputs) crucial for success and relate them, by means of equations, to design variables and outputs.
- Question design personnel to provide subjective probability distribution for design variables.
- Estimate the probability of obtaining the desired performance by appropriate techniques (e.g., simulation).
- Monitor changes in the probability of achieving target goals in performance.

Clovis et al. (1975) provides a step-by-step procedure for setting performance criteria. Briefly stated, performance criteria are established by having experts rate the significance of the various cost and achievement measures, calculating and applying weights to those measures and combining the set into a single performance criterion. Klein (undated) established four areas of criteria: a) accuracy, number of hits, hit probability, first round hit probability, engagement probability, distribution of near misses b)sustainability (number of hits per pound and per basic load); c) responsiveness (time to fire first round, time to first hit, time between rounds, time to shelf fire; and d) reliability (number of malfunctions, number of rounds between malfunctions and time to clear malfunctions.

Indices of probability of accomplishment and performance time estimates were utilized by Mitchell et al. (August 1967), and in order to convert mission specifications into a summary measure, Connelly et al. (1977) constructed a cost index function which identified deviation from the desired end state and variable rates of change, control actions and deviations from referenced trajectories occurring along the solution path.

In Cunningham's (1978) study of the adequacy of organizational resources, each organizational resource (vehicles, equipment and personnel) was assigned a cost as an indication of its value, and the resource value needed for the resolution of a given problem could then be calculated.

Finally, mission relevant performance criteria developed by Self (1972) included the following:

- The number and percentage of targets that are detected. An ideal observer would detect all targets that were displayed with some minimum image quality.

- An ideal observer would make no mistakes in designating targets as such.

- Targets should be adequately recognizable at long slant range.

14. Measurement Procedures

Ideally, discussion of measurement procedures should include: specification of the technical and procedural details concerning the measurement application; the selected measures and the specification of the data that are required; where, when, and how the data are to be collected; and how quality control over the c[illegible]y collection process is to be effected. In addition, it would seem appropriate to discuss under this heading details of the site selection, equipment use, data collection forms and materials, pre-test information and contingency plans.

It appears, however, that the authors tend, in general, to report this type of information when they discuss the implementation of the test itself. Therefore, the reader will find this topic treated more completely under Section 20, Test Execution.

15. Analytic Methods

Within the context of this review, analytic methods are considered to be a planning component of the measurement and evaluation process. However, in this as well as other planning components discussed in this report, it appears that the authors reviewed tended to report what they actually did rather than describe the previously developed test plan.

There are some exceptions, however, Spencer (1967) gave some background information with regard to the means which were devised to relate changes in maintenance performance to changes in aircraft effectiveness. In that study, a simulation model was constructed to include measures of functional reliability and alternative personnel utilization and was used to establish payoffs in terms of increased aircraft utilization and cost savings which could be compared to the cost of maintenance information system improvements. Clovis et al. (1975) discussed how weighted variables are developed by having experts rank each measure and a statistical regression performed on the set of rankings. Then, another multiple regression procedure is used to combine these measures into a single index of performance for comparison with pre-set criteria. Bovaird et al. (1959), in their study, discussed methods of predicting the expected operational availability of a system at each performance level by simple mathematical means. Two types of models were considered in Phatak's (1973) study. One was the input-output stochastic linear-state variable model and the other was the optimal control model developed by Kleinman et al. (1971). Klein (undated) reported on a technique for performing a primary analysis by use of a 3x2x2 factorial experiment, and subsequent development of a linear model on which an analysis of variance can be performed.

Bond et al. (1970) discussed evaluation designs that are considered applicable to the assessment of training effectiveness. They include the classic Solomon four-group design; iterative adaptation to individual student progress; response surface designs; adaptive control models, decision theory models and

simulation models. Karush (1969) described an analytic approach which has several measurement options. These techniques include sampling measurement, trace measurement, accounting measurement, logical and playback measurement. It was also stated that the stimulus approach has several practical measurement options depending upon the environment in which the system runs.

16. Parameter Determination

This section deals with the determination of the parameters that were selected for control in the research reviewed. The term "parameters" applies to the setting of conditions for collecting data on experimental variables, such as class intervals to be compared, number of replications to be completed, test sequences and randomization patterns to be followed, and other procedural constraints. In the narrower sense, parameters are distinguished from experimental variables in that the former are set at a fixed value or level for the duration of the experiment. For example if subject age is fixed at, say, 18-25 years for all participants, then age is a parameter. If, on the other hand, the experimenter chooses to compare outcomes across different age groups, he/she may select two or more subject groups (e.g., 18-25, 34-41, 55-62 years)—and in that case, age is an experimental variable. These distinctions are not always adhered to in the literature reviewed. It was observed that although, in many cases, the parameters established in the reviewed studies were presented, there was not a great deal of discussion with regard to how they were selected or how their levels were established.

Hicks (1977) described the test parameters used in his study. Drivers who could not meet minimum performance standards were eliminated, interviewers were trained drivers were familiarized with procedures, and the test course was established to give a variety of representative tasks over appropriate terrain. The order of driving the course was counterbalanced and the same interviewee interviewed a given driver while the driver was still in his seat. Meister (1978) stated that characteristics of equipment, job, individual aptitude, skill, experience and motivation must all be defined with regard to specifications for personnel parameters. Similar parameters were described by Sauer et al. (1977). They included time to complete tasks, career field involved in the task, degree of hazard inherent in task and task clarification. Jaschen (1975) listed day and night conditions, live fire conditions, organization, doctrine, training and logistical support as the "constant variables" and terrain and weather as the "uncontrolled variables." Van Acker et al. (1968), Repperger et al. (1978), Fischl et al. (1968), Hyatt et al. (1975), Buckley et al. (1976) and Vreuls et al. (1975) were among those authors who gave brief descriptions of some of the test parameters determined appropriate for their studies, but in general, information was lacking in the reports reviewed regarding the process involved in determining parameter selection.

As noted previously, some researchers appear to use the word "parameters" when discussing the experimental variables. For example, Meiser (1967) stated that theoretically, data on "parameters" can be obtained under either experimental conditions or the actual operational environment. However, the author doubted that experimental work is able to supply the necessary information because of difficulty of data collection in the operational environment. He stated that one difficulty is setting up conditions which isolate the "parameters" of interest; the "parameters" usually exist only in interaction under operational conditions. One solution might be the identification of the operational conditions which display

combinations of "parameters" of interest. By locating and measuring different "parameter" combinations, comparing results and describing differences to variations in the "parameters" between the two combinations, the individual "parameter" effects can be isolated.

Levy (1968) states that "parameters" can be derived from either theoretical or empirical models. Theoretical models are obtained by a hypothetical deductive procedure. They provide a guide to research and a program activity which empirical models lack, but theoretical models involve assumptions which may or may not have been tested. In empirical models, on the other hand, equations are obtained by curve fittings. Another issue of concern in model formulation noted by the author is the question of "what kind of man is being modeled." Is it an individual with individualistic parameters or is it the "average man?"

17. Apparatus for Testing

This topic should describe all of the equipment used to perform the measurement and evaluation study. For example, Obermayer et al. (1974) provided information in four specific areas:

a. Monitoring and data collection equipment

b. Equipment being tested

c. Environmental facilities used for the test

d. Data processing equipment

In the literature reviewed not all of the authors provided this information and often descriptions of the equipment were brief and incomplete. In addition to Obermayer et al. (1974), U.S. Army Infantry School (1976), U.S. Army Test and Evaluation Command (1971), McKendry et al. (1964), Goldbeck et al. (1971), featherstone et al. (1975) and Rhoads (1970) were among those authors who presented reasonably comprehensive discussions. However, in the main, the equipment used was generally described in vague terms.

Monitoring and data collection equipment mentioned in the U.S. Army Test and Evaluation Command (1971) report included linear measuring devices (tape measures, rules, etc.); weighing devices (scales and balances); sensors (temperature and pressures); and meters (light, sound and vibrometers). Other studies (U.S. Army Test and Evaluation Command (1970), Dunlap and Associates, Inc., (1966), etc.) utilized photographs, motion pictures, videotapes, tape recorders and questionnaires.

The equipment being tested, of course, is directly related to the objectives of the study. For example, Featherstone et al. (1975) reported a comparison study of a .45 caliber automatic pistol and a 10-38 caliber revolver, and Klein et al. (1969) tested two different automatic rifles. Some of the equipment or systems

being evaluated, required highly sophisticated simulator systems. For example, Buckley et al. (1976) utilized an air traffic control simulator and Vreuls (1975) employed an automated instrument flight maneuver trainer in this study.

With regard to the environmental facilities, in some cases, the evaluation took place in a real-life environment and, in other instances, simulated facilities were utilized as in the study by Klein et al. (1969) where the automatic rifles were tested in a simulated combat-firing facility.

The U.S. Army Test and Evaluation Command (1971) required the use of an acoustical chamber, and Connelly et al. (1977) used a surface ship bridge console system and a CRT. Finally, the equipment used in the analysis of the data depended again upon the complexity of the evaluation. It ranged from analyses performed by the use of two ordinary tables (Mills et al. 1974) to data processing facilities for detailed analysis and evaluation (Obermayer et al. 1974).

It would appear that the four equipment topics suggested by Obermayer et al. (1974), i.e., monitoring and data collection, test equipment, environmental facilities and data processing, are appropriate and useful in reporting on systems measurement and evaluation efforts. The review of the literature revealed, however, that in many instances information is lacking in one or more of the above mentioned areas.

18. Personnel for Testing

"Personnel for testing" includes the subjects who are being tested and the experimenters (or testers) who conduct the research. The numbers of persons and their relevant attributes are both of interest here. In this review, few authors were found to document their rationale for selecting a particular sample size to test subjects, and its appears that in most cases, sample sizes were determined by matters of convenience like time, money and availability. In addition, little attention appears to have been given to defining the requirements for test personnel. Fineberg et al. (undated) and Siegal et al. (1970) did describe the type of personnel they used as testers but, generally, information was lacking in this area.

The sample size for subjects who participate in testing should depend upon the experimental design and statistical analysis techniques employed to meet the objectives of the study. In some instances, small sample sizes were obviously considered appropriate by the researcher (Wellman et al. 1972). Sample sizes ranged from as few as three to an entire military unit. Wellman et al. used only 35 subjects in their study to develop performance measurement standards. Hansen et al. (1977), on the other hand, utilized 445 airmen in an Air Force technical training validation study. The characteristics of the participants should vary according to the type and objective of the study. Many factors can enter into the choice of participants: age sex, educational background, skill level in the appropriate discipline, etc. Most of the studies reviewed were of a military nature and the populations sampled ranged from highly qualified technical personnel such as pilots, to unskilled enlisted men. In some of the studies, specialized personnel were required such as the research reported by Fineberg et al. (undated) who utilized Army helicopter pilots with nap-of-the-earth experience. Goldbeck et al. (1971) required that his college student subjects met certain physical and scholastic criteria and Hill et al. (1974) selected his subjects based on their flying experience.

In summary, subject selection and sample size is entirely dependent upon the types of research conducted and, in the literature reviewed, a wide range of requirements were reported. However, there was little documentation of the rationale and determinants used for defining sample size, subject characteristics or test personnel.

19. Test Plans

In this review, test plans are considered to be the summarizing step of the anticipated measurement process, in which the analyst's decisions are formally documented for review, reconsideration and revision for final implementation. In the final reports which were reviewed, many researchers only discussed the actual implementation of the test itself. When test plans were presented, they typically provided only a brief outline followed by a more complete description of the test execution. Several were a little more detailed. For example, Berson et al. (1976) describes the test activities in his research to obtain and analyze human performance data. Five steps are presented in his test plan: 1) test administration (includes milestone development, manpower specification and budget preparation); 2) task group identification and task analysis (defines behavioral requirements, performance standards and specific functions of the system); 3) test planning and design (includes test objectives, selection and design of test equipment, test environment and test personnel selection); 4) test execution (describes procedures for conducting the activity and includes a pre-test; and 5) data analysis techniques and the determination of the appropriate technique applicable to the data collected.

In another research example, Performance Measurement Associates, Inc. (1978) describes a test approach for performance measurement devlopment. The author presents a method of constructing a task component measure (TCM) that relates the quality of performance of each task component to the summary performance measures selected using a computer processor. In a third case, designed to test the feasibility of developing and utilizing personnel performance effectiveness measures for man/machine function allocation decisions, Willis (1967) suggested a research plan involving four steps: 1)the selection of parameters and observation on a simulated system; 2) the testing and refinement of parameters; 3) the determination of how the methodology might be implemented; and 4) the development of an automated system for handling data. Brown (1977) utilized the USA ARENBD Test Design Plan for Field Evaluation of the M48A5 Tank Product Improvements. In this plan, for each product item, detailed descriptions were presented of test procedures including objective, method, analysis and results. Andrews (1977) presented a detailed test plan for the initial test and evaluation of a radar set. The operational test consisted of testing in all primary and secondary modes. Varying flight profiles were used to assess detection ability and tests were designed to be conducted under all weather conditions.

Foley (1975) discussed the test factors which were considered in his study concerning the technical proficiency in maintenance activities. They included the identification and classification of the tasks to be measured. Consideration was given to the hierarchical relationships of maintenance tasks, and the most effective order of measurement and the ease of test administration. Swink et al. (1978) defined requirements for a performance measurement systems for aircrews. The first phase developed candidate performance measures from documentation, and interviews with operationally qualified aircrews. In addition, a special purpose evaluation sortie for the simulator was developed. In Phase 2, several alternative

configurations were designed to meet performance measurement requirements by reviewing the existing system and documents and by interviewing techniques. In the last phase, the functional and engineering requirements for the performance measurement systems were described.

Another implementation plan, by Obermayer et al. (1974), was based on the Air Force Systems Manual AFSCM 375-5. Five major steps were recommended: 1) selection of system integration contractor; 2) completion of preliminary detailed system design; 3) selection of final system design with testing; 4) procurement of hardware; and 5) completion of final system tests. Rasch (1973) cited the following elements as playing a determining role in the implementation of a technical performance measurement (TPM) program: 1) parameter selection and documentation of detail; 2) construction of TPM models; 3) profiling parameters; 4) planning the TPM; 5) assessing organizational participation; and 6) preparation of reports, data analysis and predictions.

Finally, Klein (undated) reported the components of a test plan which was comprised of selection of subjects, determination of sample size, weapon assignment, training of subjects, scheduling, test facility determinants, test implementation and data analysis.

In conclusion, it should be noted that much of the research reported lacked adequate descriptions of the test plan utilized and, in other work, descriptions of the test plans were sketchy. Although reporting was generally inconsistent, this review represents an effects to describe in an uniform manner some of the techniques discussed by the authors on planning for the measurement process.

20. *Test Execution*

Test execution is what most authors seem to be reporting as what they did during the execution of the test. We would have expected to have been informed about the degree of conformance with, or departure from the test plan. However, typical project reports that have been reviewed rarely make reference to the test plans but simply report how the tests were implemented. A large volume of material was obtained from this topical area and it is not possible to include all of the test reports reviewed. Therefore, discussion here is limited to a brief description of a sampling of the work which has been performed.

In a study to determine the effects of system and environmental factors upon pilot performance in an advanced simulator for pilot training , Irish et al. (1977) described the test procedures in which each subject flew one profile 72 times and the other 27 times during the course of the study. The profiles were randomly ordered for all subjects. Each session was begun with instructions provided by a computer driven word generator. Each maneuver was begun on command and completed when selected cirteria were satisfied. At the completion of each maneuver within the profile, the console operator entered comments on any system malfunction or errors experienced during the maneuver. All other data values were recorded by an ASPT computer.

A methodolofy and criteria were established by Turner et al. (1972) to assess a system's capability using system-level measures of effectiveness. A set of MOE's was established for both the airborne warning control system and the tactical mission levels. To obtain a standard for comparison, the scenario under consideration was analyzed both with and without the AWACS. The incremental differences

in tactical mission MOEs combined with the AWACS system MOEs provided insight into the effectiveness of AWACS.

Spyker et al. (1971) describe their approach to developing a measurement workload index and physiological workload index based on a pilot's physiological response to a simulated tracking task. The procedural steps described include validation of a sensitive nonloading secondary task, collection of physiological and performance data, extraction of the potentially meaningful data, normalization of the features and selection of the "best" subset, computation of the workload index and the best linear predictions from the subset and, finally, validation of the predictor. Three direct measures were obtained from these efforts—miss rate, response time, and an evaluation of task difficulty. In Hicks' (June 1972) evaluation of vehicles in operational field tests, a human factors vehicle evaluation instrument methodology was utilized. A driver, upon completion of a test drive was interviewed to obtain ratings on a six-point rating scale. In addition, the drivers were required to rate the relative importance of 85 vehicle characteristics.

In an aircrew oxygen system development study (Kiraly et al. 1970), animal and human tests were conducted. In the first test the animals received a single acute exposure of 3.5 hours duration and a chronic exposure of 5.5 hours/day for ten consecutive days of rebreather gases. The animals were sacrificed and lab examinations conducted on lung tissue. In the human tests, two series were conducted. In the first test, subjects experienced the system operated with and without safety pressure to determine comfort levels and possible physiological damage to respiratory systems. The second test was to determine relative comfort of alternative equipment. In both tests, carbon dioxide levels and oxygen levels were monitored. Mask leakage measurements were also made in relation to the employment of safety pressure and comfort levels.

In an IOT&E of a radar imagery recorder (Chasteen, 1975), the following test was executed. Six missions of eleven sorties were flown under controlled test conditions; known checkpoints and offset aiming points were used by the navigators to provide independent evaluation of the system. Each sortie was flown at preselected altitudes ranging from 500 feet AGL to 25,000 feet MSL. Routes were preselected and enroute position coordinates were recorded on data forms along with intensity/gain used. A camera/periscope assembly recorded the radar display and auxiliary data throughout the mission. Debriefing meetings, attended by various specialists, included review and analysis of the recorded imagery; comments and recommendations were solicited from the attendee relative to his area of expertise. Questionnaires were also completed by the navigators who participated in the study.

In the development of automated GAT-1 performance measures, two experiments were conducted by Hill et al. (1974). A warm-up period was allowed to familiarize pilots with the GAT-1 and its flight characteristics. This warm-up period varied depending upon the skills of the pilots. In experiment 1, the major tasks were:

- Roll and pitch tracking
- Roll and pitch tracking with power changes
- Flight profile

- ILS landing approach

In experiment 2, the following tasks were added:

- Roll tracking
- Roll, pitch and yaw tracking
- Reduced bandwidth roll tracking
- Reduced competence roll tracking
- Ground reference turning maneuver
- Attitude position tracking

Repperger et al. (1978) reported on an experiment to evaluate parameter changes on the human operator under thermal stress. These subjects were exposed for one hour to a simulated heat-soaked aircraft environment. They performed a single dimension compensatory tracking task for 5 minutes duration, separated by 5 minutes of rest. The tasks represented flying a very stable aircraft under vertical wind buffet. Each subject participated in six experimental conditions, three control runs and three exposures to the heat-loading environment. During the experiment, the subject maintained one of three conditions of water-electrolyte balance. He either drank nothing or replaced weight losses with water or Na Cl solution. The subject urinated and blood samples were drawn periodically. Mean skin temperature, rectal temperature, weight loss, heart rate, air temperature, water temperature, and humidity were recorded along with tracking performance parameters.

Greening (1968) validated the model used in this study by comparing the model's predictions for selected targets with the results obtained experimentally in which a number of observers viewed motion picture presentations of a flight over the target. Mumford et al. (1961) developed performance criteria for turret mechanics. In this study, information was collected on the task at the organizational level by studying job descriptions and interviewing consultants knowledgeable in the field. Tasks selected for the study reflected expert judgment and exercises and tests were developed and administered to subjects. A scoring system was developed which was able to distinguish degrees of adequacy or inadequacy of performance. Erickson (1968) reported on a field experiment conducted to validate a mathematical model of the visual detection process. Pilot observations were recorded of a non-dimensional visual search. Farina et al. (1971) utilized task characteristics rating scales which were subjected to multiple regression analysis to establish the extent to which they were performance related.

Hansen et al. (1977) administered reading aptitude tests to provide predictor performance scores in the development of a flexilevel adaptive testing paradigm. The training of student pilots on the automated adaptive flight training system was compared by Grunzke (1978) with operational crews who received experience in flying the F-4B WSTSH15. Crews flew and were scored on different types of air-to-air intercepts that were programmed into the training device. Chasteen (1975) evaluated a radar imagery recorder. Navigators flew sorties under controlled test conditions to provide an independent evaluation of the radar system.

The report by Erickson (1968) described a field experiment conducted to validate a mathematical model of the visual detection process. All observations were made by pilots flying A-4 aircraft above a bulldozed strip in the desert. Ground targets were a Sherman tank and a radar van without the radar dish/antenna. Thus, the visual search was in one dimension only; the model was not capable of handling two-dimensional search. Flights were conducted at altitudes of 1000, 2500, and 4000 feet, at indicated airspeeds of 275, 270, and 265 knots, respectively.

In an evaluation of operator loading in man-machine systems, Siegel et al. (undated) conducted a test in which the experimental subjects tracked continuously for eight hours between 8:00 a.m. and 4:00 p.m. No breaks were allowed. Samples of their performance were recorded for the last five minutes of each hour. The subjects were unaware how much, if any, of their performance was being recorded. Transcription of the data involved measuring the displacement of the input and output signals as recorded on the ink writing oscillograph.

Taylor et al. (1977) performed a human factors engineering study of two ball port designs for an infantry fighting vehicle. In this study, each subject was trained to install and remove the weapon on each configuration. The seat chosen for the experiment was at the worst possible angle for the tasks. Each subject performed six trials in removing and installing the weapon in each design configuration. Time measurements were obtained by means of a stopwatch.

In a study conducted by Phatak (1973), subjects performed the tasks at sea level followed by the same task at a simulated altitude of either 12,000 or 20,000 feet. Each run of this experiment consisted of two tracking periods. Each period was preceded by one minute of pre-breathing at the indicated altitude followed by one minute of tracking. Randomization of the order of presentation of the simulated altitudes and tasks to the six subjects was done in order to minimize the effects of learning and anticipation of experimental factors.

Featherstone et al. (1975) attempted to determine measures of effectiveness for the handling characteristics of small arms. The weapon, task sequence, and subject factor levels were set prior to conduct of the experiment. Four task sequences were selected. Subjects were briefed prior to the experiment and received written instructions. Practice on both types of weapons was permitted followed by actual firing in the task sequence assigned. At the end of the firing, the subject filled out an information sheet giving personal data and weapon evaluation.

The roles of vision and audition in truck and bus driving were evaluated by Henderson et al. (1973). To evaluate experimentally both the results of the analytical effort and the test device, the entire battery of visual and auditory tests were administered to the subjects along with a questionnaire to derive biographical and driving pattern data. Driving record information was obtained from company files for each driver tested, including total number of accidents on file, number of "responsible" accidents on file, number of months covered by the files, and total number of accidents and "responsible" accidents for the last 36 months.

In a previous study, a Technical Behavior Checklist was developed for four naval ratings. This checklist was a detailed comprehensive checklist of the tasks performed in that rating. For this study (Siegal et al. 1961), a supervisor was asked to indicate the proficiency level of the man he was rating in terms of how much supervision and the number of checkouts required.

Fineberg et al. (undated) assessed navigation performance of Army aviators under nap-of-the-earth conditions. The navigators were assigned missions in which designated landing zones had to be found for simulated medical evacuations or supply deliveries. All 35 aviators navigated at least six NOE routes ranging from 23 to 25 kilometers (km) in length. Twenty-eight of the aviators were also tested on aircraft control and the performance of various NOE maneuvers. Harry (1975) reported on a study to determine utilization experience of public services. Two cities (St. Petersburg, FL. and Nashville, TN.) were chosen as the experimental sites for this study. Data were collected by appropriately designed surveys of the user or providers of services. A computer program for organizing and analyzing the data was developed.

In a study of the new SAINT concepts and the SAINT II simulation program, Wortman et al. (1975) described a test simulation of aircraft refueling In the test, the receiver and tanker are initially flying at the same velocities. Perturbations of the tanker's velocity are incorporated in the model and represent environmental disturbances (turbulence). The objective of this simulation was to determine how well the receiver pilot is able to maintain his refueling position in the face of these disturbances and the prescribed control strategy.

Companion et al. (1977), in an application of task theory to task analysis, executed a test whereby problems performed on desk and pocket calculators were developed so as to represent theoretical tasks. Ten subjects were instructed in the theoretical concepts, and were then provided a partial operational analysis of the task problem. They were then required to complete the operational task analysis and to transform it into a theoretical task analysis.

21. Data Analysis

The statistical techniques employed in the data analysis of the studies reviewed varied according to the complexity and nature of the work performed. These techniques ranged from fairly simple mathematical calculations such as determining averages, to analyses using sophisticated computerized equipment and programs. It should be noted that few researchers discussed their rationale for their choice of the particular statistical tools used in these analyses. Listed below are some of the kinds of techniques reported in the literature. They have been grouped into four areas—descriptive statistics, measures of association, measures of statistical dependence and general systems analysis.

a. Descriptive Statistics (Blanchard et al. 1969; Bloom et al. 1979; Buckley et al. 1976; Dunlap and Associates, Inc. 1966; Farina et al. 1971; Grunzke, 1978; Helm, 1976; Henderson et al. 1973; Hill et al. 1974; Hyatt et al. 1975; Klein et al. 1969; Lindsey, 1974; Mills et al. 1974; Siegel et al. 1974; Siegel et al. 1970; Siegel et al. 1961; Siegel, undated; Timson, 1968; Waag et al. 1975; Weapon System Effectiveness Industry Advisory Committee, 1965, Vol. 1)

- Means, modes, medians
- Standard deviation and ranges
- Error values and scoring techniques
- Frequency analysis

- Matrix displays
- Histograms
- Performance measure scores
- Graphical analysis and mapping
- Critical path analysis
- Scaling

b. Measures of Association (Buckley et al. 1976; Cunningham et al. 1965; Dunlap and Associates, Inc., 1966; Henderson et al. 1973; Kribs et al. 1977; Meister, 1978; Melching, 1968; Sheldon et al. 1967; U.S. Army Infantry Board, 1971)

- Regression analysis
- Factor analysis
- Kendall's coefficient of concordance

c. Measures of Statistical Dependence (Chop, 1972; Cunningham, 1978; Dunlap and Associates, Inc., 1966; Goldbeck et al. 1971; Grunzke, 1978; Hicks, October 1977; Hicks, June 1977; Highsmith, 1976; Mills et al. 1974; Rhoads, 1970; Repperger et al. 1978)

- Analysis of variance
- T ratios
- Chi-square tests
- Probability analysis
- Post-hoc multiple comparisons
- Duncan's multiple range test

d. General Systems Analysis (Blanchard et al. 1969; Brokenburr, 1978; Helm, 1976; Hill et al. 1974; LTV Aerospace Corporation, 1973; McDonnell Douglas Astronautics Company-Eastern Division, September 1969, Book 1; Sauer et al. 1977; Thurmond, undated; Wellman et al. 1972)

- Mathematical modelling
- Linear and non linear modelling
- Simulation modelling
- Critical incident techniques
- Subjective judgments by experts (e.g., Delphi technique)

22. Findings and Interpretation

Some authors present their findings in ways which are more useful than others. The presentation often depends on the purpose of the study and is

framed in terms of the study's hypotheses. For example, Klein et al. (1969) reported that in a comparison test of two weapons, one had more potential than the other. Sometimes the findings were not anticipated in the original formulation of the objective so they are presented as an unexpected product of the study. Other studies are analytical in nature and develop a line of reasoning to prove or demonstrate a point of view. In some cases, the findings constitute the basis for further research as in the study performed by Finley et al. (1976) which presented a preliminary model of a systems taxonomy model consisting of three major levels: 1) systems objectives, 2) system functional purpose and 3) system characteristics. These three levels are further defined by their relationship to the nominal versus relative levels of measurement. This particular report served as one point of departure for the System Development and Evaluation Technology being conducted by Dunlap and Associates, Inc., of which this state of the art report is a subtask.

Some authors feel that their work has produced definitive conclusions, for example, Goldbeck's (1971) major finding of his study concluded that the optimum application of the Sequencing Technique is the most powerful tool available to the control panel designer, and Akashi et al. (undated) showed that the performance of operation can be represented mathematically. Connelly et al. (1969) stated that although much effort has been devoted to the problem of improving human reliability data, there has been little conceptualization of the overall problem and a lack of development in the state of the art of quantifying human performance effectiveness. The results of other studies are often less definitive and the authors present their findings in a more subjective or judgmental manner. In the area of task analysis, Companion et al. (1977) noted that the results of this study appeared to indicate that, with very little training, people can comprehend the concepts and be at least as proficient in theoretical analysis as they are at describing actual operations. Therefore, operational task descriptions or task analysis can be translated correctly into the tasks of the theory by minimally trained observers. Farina et al. (1971) found that it appears possible to describe tasks in terms of task-characteristic language which is relatively free of the subjective and indirect descriptions found in other systems and, further, that task characteristics may represent correlates of performance. Siegel et al. (undated) suggested that use of the spectral analytic techniques possesses considerable potential as an on-line assessment of operator status in man-machine systems involving perceptual motor behavior. Finally, in model development, Levy (1968) stated that the precision or accuracy required of a model is generally regarded to be a function of the stage of the system life cycle in which the model is being used. However, the author believes that the same levels of precision are required in the initial stages of design as in later applications. In order to be truly useful, applied models must consider the relevant interactions of design parameters with difficulty of conditions and sufficient degree of accuracy.

In summary, it can be seen that findings and interpretations are reported generally in terms of the original hypotheses, other facts which can be obtained from this type of study conducted, or in support of a position being promulgated by the authors.

23. Conclusions and Recommendations

Conclusions and recommendations are often presented as a summary of the findings and interpretations with perhaps additional emphasis on the implications of the research and the identification of further research needs. One would also

expect that some definitive statements would be found regarding system effectiveness and performance. The conclusions and recommendations might even contain recommendations for continuation of the research to verify the results. The hypotheses might be restated when appearing as a conclusion, and there could also be a discussion of any limitations or restrictions identified with (or by) the work.

As well as reporting definitive results, many researchers reviewed during the preparation of this report concluded that their techniques could be useful to others in the field. For example, Ellis (1970) recommended that the techniques used in this study be included in the repertoire of those considering man/machine interface analysis. Hutchins (1974) discussed the responsibilities of human factors engineers in defining system specifications. Many researchers also identified future research needs—Helm (1975) was able to narrow the problem area of human factors design deficiencies and recommended that future efforts be devoted to two particular areas of concern. Geddie (1976) felt that long-range benefits would result from his approach to the development of performance based criteria and Hankanson (1967) suggested that his model, although presented in simple terms, has use of handling systems and tests of a highly complex nature. Haight (1971) noted the need for validation of his results and Dunlap et al. (1967) discussed the limitations imposed on their study by the lack of necessary equipment.

In summary the factors mentioned above would appear to have been covered appropriately in most of the literature reviewed. However, the conclusions of a study can only be as powerful as the study design itself, which is sometimes inadequate for the intended evaluation purposes.

IV. MEASUREMENT LIMITATIONS AND PROBLEM AREAS

The problems associated with the assessment of manned systems seem more of not knowing what performance to measure or which methods to use rather than not knowing how to measure. There is little that is uniform or systematic about the various approaches to manned system measurement. The system itself is often overlooked as an important source of variables affecting the selection and application of performance measures. The system components sometimes have been studied and measured out of context, in isolation, as if there were no need to consider the systems in which they are imbedded. On occasion, evaluators and designers have studied and "improved" performance of operators and crews in a nonsystem-specific context, only to find, when the operator/crew was returned to the system, that real system performance had not benefited at all. The system's unique variables can profoundly affect the personnel subsystem's performance. One factor that probably exacerbates measurement and evaluation problems is the wide variety of manned systems, their tremendous diversity of purposes, and their many variations of size and complexity. This makes it difficult to view systems as entities that belong to the same universe and that form important populations and subpopulations within that universe.

The remainder of this section briefly describes some of the limitations and problem areas in the measurement of various individual systems and of systems in general, as noted by researchers in the literature reviewed.

Severe limitations on system measurement were noted by several authors (Blanchard et al. 1969; Clovis et al. 1975; Kelley, 1968; Levy, 1968; Meister, 1968; Pew et al. 1977; Rigby, 1967; and Ultrasystems, Inc., 1972). One concern, expressed by Blanchard et al. (1969), is the lack of valid human performance data, a problem which can seriously limit the utility of an evaluation model. Subjectively derived performance data continues to be given prime emphasis and it is felt that this is not likely to change soon. Clovis et al. (1975) suggested that in dealing with the limitations of their study, a cross validation effort be conducted to test the efficiency of the regression equations used in calculating the index of performance. Also, it is recommended that situational exercises be used to validate and to provide practical application guidelines. Kelley (1968) makes the point that human performance is not linear and may be poorly represented by linear control-theory models except for fairly simple or restricted tasks. Also, human control is exercised not on the basis of present error, but rather on the basis of future (anticipated) error.

A need for validation of man-machine models is noted by Levy (1968). It was suggested that a research design for developing and validating applied models be undertaken. This design would call for the collection of performance data and input data in field situations with the input data recorded for use in laboratory studies aimed at model development. The models, in turn, would be validated by comparing their outputs with the pre-collected field performance data. Meister (1968) discusses the human reliability model primary as a means of illustrating certain characteristics of behavioral models in general and certain characteristics of model makers themselves. In the author's view, a model is effective to the extent that it helps to either gather data and/or to explain those data. He states that any behavioral model which is not concerned with real-world data (as opposed to laboratory data) is not useful. However, he observes that behavioral models

characteristically employ laboratory data and have ignored or have been unable to handle natural event data. The author asserts that the human reliability model's assumptions derive from the unsystematic manner in which the model's input data were secured and that, at least in part, these assumptions demonstrably are not in accord with empirical reality. Pew et al. (1977) noted that, for the most part, human information processing models deal with the average performance of well-motivated, high practiced individuals under relatively ideal conditions. There are many hypotheses but few data and virtually no models in the information processing literature on how human performance capacities change under stress, reduced motivation, or before practice has stabilized performance. Rigby (1967) asserts that the development of an accurate data base of human error rates is impeded by several factors—accidents and mission failures resulting from human error are not reported as regularly or as accurately as equipment failures, and that there is a lack of standardization in terminology, manner of development and level of reporting.

Finally, Ultrasystems, Inc. (1972) presents 12 areas of limitations on system measurement. First of all, it is stated that the criterion for success is seldom explicitly stated and that there exists more than one way of defining a mission as well as more than one way of quantifying how well the criteria for success are met. It is noted that the rationale for MOE selection is not always presented and, in general, the MOEs used are those that are readily obtained via model development. Very seldom, when more than one MOE is identified, is a ranking of importance performed or a combined measure developed and used. Expected value type MOEs are most prevalent in force level studies, whereas probability type MOEs are most often found in subsystem level studies. With regard to independent variables, it is felt that over twice as many occur in the friendly force category than in the threat and target categories combined. In addition, as the study level increases from subsystem to system to force level, the percentage of independent variables in the friendly force category decreases and the friendly force interaction with threat or target category increases. It is noted that there are cases where the variables selected for model formulation are not readily (if at all) measurable in the real world. Physical environment aspects appear to be generally ignored or casually treated in effectiveness studies and, finally, it is not easy to compare similar effectiveness studies.

In summary, it appears that limitations of major concern to those developing models are the lack of valid human performance data (resulting in part from the absence of information on performance under "real-world" conditions), lack of standardization in development and reporting of data, and the need to validate man-machine models with field performance. In addition, limitations in system measurement are reported to exist in the areas of defining a mission and quantifying its success, lack of rationale in MOE selection and the selection of variables which are measurable in the real world.

V. PRIORITIES FOR MEASUREMENT IMPROVEMENT

There was some discussion in the research reviewed of priorities for measurement improvement. In some cases, recommendations were limited to the system of specific interest to the author, but generally the recommendations for future research were directed toward verification of the research just completed. Several authors, however, made recommendations for future research which would have broader application to system measurement improvement.

Willis (1967) noted the absence of a body of quantitative evidence about the performance effectiveness of personnel in present systems. It was suggested that, as a first step, a data bank on personnel performance be developed which would select samples of personnel performance which could be generalized to entire classes of populations. In a study designed to evaluate a military system, Dunlap and Associates, Inc. (1966) noted that they had identified promising directions for further study in the areas of team data analysis and performance testing. They recommended that large quantities of data from multiple trials be methodically built into a data base for each variable, team member and subtask of a standard test. In addition, they recommended objective field monitoring techniques such as video recordings should be utilized to provide standard structured coverage by separate variables and subtasks.

Recommendations for further research and development of large-scale system modeling efforts included the development of a test to evaluate alternative model formulation of common task environments and to conduct empirical validation studies to compare model prediction with actual human performance. Pew et al. (1977) recommended methodological research on the implications of combining subtasks or information processing component models on the aggregate system performance. It was felt that further research should be conducted on the validation of large-scale simulation models and guidelines should be developed for the acceptable number of free parameters in useful predictive models. Williams (1967) addressed the problem of estimating conditional probabilities of dependent task steps. He said that these problems can only be solved by developing transition models that make the transformation from marginal probabilities of the data store to the conditional probabilities of the dependent relations. He noted that two major problems must be solved before there will be significant progress in developing transition models. These are: 1) the identification of factors responsible for dependent relationships among task steps, and 2) determination of the effects of dependent relationships.

With regard to the measurement of the proficiency of maintenance personnel, Gustafson (1967) recommended several areas for future research. The major goal of this research would be aimed at developing specifications of proficiency measures for inclusion in weapon system development contracts. This research would include refining principles and techniques for assessing individual performance in trouble-shooting and other complex tasks, continued research to improve maintenance records and supervisor ratings as job performance criteria, and the development of a practical procedural handbook which can be followed in assessing performance capabilities. Companion et al. (1977) felt that the approach in their study which applied task theory to task analysis evaluation of validity to reliability using simple tasks, should be extended to the evaluation of more complex tasks.

Brokenburr (1978) concluded that not only is future research needed to verify results of his study but that learning curves should be developed for specific teams (such as rifle squads). Knowles et al. (1969) suggested several approaches for studying system design which relate to the evaluation of equipment-oriented tasks. Pritsker et al. (1974) suggested verification of the factors and relations included in the characterization of task performance. In addition, it was recommended that new concepts be developed in order to model tasks that require continuous monitoring, queuing and resource allocation, and that the treatment of task type and the method by which operations are assigned tasks be extended.

In summary, therefore, recommendations for future efforts for measurement improvement included the development of quantitative data on personnel performance, further development of objective field monitoring techniques, the development of transition models as well as research and development of large-scale system modeling efforts.

VI. CONCLUSION

This state of the art assessment of manned system measurement reflects the review and abstracting of over 250 relevant technical documents. The documents cited under each category of interest are representative of the literature generally, and do not purport to be the total of all reviewed documents that contained any relevant information. Those cited in each category of interest, however, are believed to describe the key published concepts and recommendations that define the state of the art today.

This report employed a topic outline compatible with the overall measurement model being developed under the present contract. Nevertheless, it is believed that the model is sufficiently representative and comprehensive so that all significant comments and authors have a place in its structure. Consequently, the appropriate state of the art information is contained here, though its arrangement or form may vary from where different models or working outlines would have it.

One of the important uses of this review is the identification of current measurement capabilities and limitations, so that requirements and priorities for the improvement of system-oriented measurement can be delineated. In this review, it became apparent, for example, that measurement models need to be further developed, supported with appropriate human performance data, refined through more consistent and comprehensive applications, and validated by independent corroborations of some kind. Furthermore, the general sense of impracticality and the need for simplifying assumptions in some cases, strongly suggests a requirement for improving the "efficiency" of measurement models by reducing the magnitude of effort required, while remaining true to the real world of the system under assessment. This latter need for procedure magnitude reduction could be accomplished in a stepwise fashion by an overall direct effort, supported by individual limited efforts for the clarification and simplification of specific concepts and the modification of analytic approaches. One of those approaches, for example, could be the introduction of computer-aided procedures employing carefully developed taxonomies and checklists.

It is envisioned that much time, effort, and money can be saved, irrelevant measurements can be avoided, and meaningfulness can be enhanced by making the kinds of improvements noted above. Ultimately, these improvement efforts could make the difference between an oversized, difficult-to-use measurement and evaluation procedure with limited acceptance and few users and a clearly established, easy-to-use procedure with wide acceptance and many users.

APPENDIX

Bibliography

AC Spark Plug Division, General Motors Corporation. Inertial guidance system 107A-2--Category II, personnel subsystem test and evaluation (PSTE) and maintenance, logistics, reliability and readiness (MLRR) test and evaluation--Objective achievement status report. (AF 04 (694)-177 AFBM Exhibit 60-20A) Milwaukee, WI, 15 February 1964. (AD-829 749).

Akashi, H. & Mahmood, S. Performance of human operators under various system parameters. (NASA CR-6725). National Aeronautics and Space Administration, (undated).

Analytics, Inc. Measuring the performance of operational decision aids (Final Report 1161-B). Willow Grove, PA: Analytics, Inc., April 1976. (AD-A024 795).

Anderson, R. Measures of aircraft effectiveness (AOS-TR-73-5). Kirkland AFB, NM: Office of the Assistant for Study Support (OAS), May 1973. (AD-913 306).

Anderson, W.H. Development of performance measures for organizational level aviation maintenance managers. NTIS Weekly Government Abstracts, October 17, 1977. (Abstract).

Andrews, L.B. Detailed test plan (DTP) for the initial operational test and evaluation (IOT&E) (OT-II) of the AN/TPS-59 radar set (C 0041-0-07-7). Quantico, VA: Commanding General, Marine Corps Development and Education Command, August, 1977. (AD-B020 848L).

Baker, J.D. Quantitative modelling of human performance in information systems. Ergonomics, 1970, 13, (6), 645-664.

Basso, G.L. A methodology for measurement of vehicle parameters used in dynamic studies. Ottawa, Canada: National Aeronautical Establishment, July 1973.

Beau, J.F. Management of the human element in the physics of failure. Paper presented at The Third Annual Symposium on The Physics of Failure in Electronics, Chicago, IL, September 29, 1964. (AD-812 593).

Berson, B.L. & Crooks, W.H. Guide for obtaining and analyzing human performance data in a materiel development project (Tech. Memo. 29-76) Woodland Hills, CA: Perceptronics, Inc., September 1976.

Blanchard, R.E. & Smith, R.L. Field test of a technique for establishing personnel performance requirements. 1969 Annals of Assurance Sciences, 272-277. New York: Gordon and Breach Science Publishers, July 1969.

Bloom, R.F., Pepler, R.D., Schimenz, M.V., & Lenzycki, H.P. IFV/CFV personnel selection analysis (ED 79-5). Darien, CT: Dunlap and Associates, Inc., July 1979.

Boehn, B.W. Computer systems analysis methodology: Studies in measuring, evaluating, and simulating computer systems (R-520-NASA). Santa Monica: CA: The RAND Corporation, September 1970.

Bohdanovych, V.I., Kulyk, V.T., & Uyomov, A.I. (Development and practical application of systems analysis.) Filosofs'ka Dumka, 1973, 1, 10-16. (Arlington, VA: Joint Publications Research Service, 1973. JPRS 58935).

Bond, N.A., & Rigney, J.W. Measurement of training outcomes (Tech. Rep. 66). Los Angeles, CA: University of Southern California, Behavioral Technology Laboratories, June 1970. (AD-711 302).

Bovaird, R.L. & Zagor, H.I. A systems approach to predicting and measuring Polaris fire control system operational availability (RM 59TMP-57). Santa Barbara, CA: General Electric Company, Technical Military Planning Operation, December 1959. (AD-901 773).

Boycan, G.G. & Warnick, W.L. Training requirements for the armor crewman and reconnaissance specialist advanced individual training programs (HumRRO-CR-D2-72-7). Alexandria, VA: Human Resources Research Organization, November 1972. (AD-759 569)

Breaux, R. Training characteristics of the automated adaptive ground controlled approach radar controller training system (GCA-CTS) (NAVTRAEQUIPC TN-52). Orlando, FL: Naval Training Device Center, July 1976.

Brokenburr, J.L. Learning curves and their applicability to unit training levels in operational testing (Master's Thesis). Georgia Institute of Technology, 1976.

Brown, D. Statistical guide for CDEC experimental design (Rev. Preliminary Draft). BDM Scientific Support Laboratory, October 1976.

Brown, J.D. Field evaluation of M48A5 tank product improvements. Fort Knox, KY: U.S. Armor and Engineer Board, 14 January 1977. (AD-B016 139).

Buckley, E.P., Goldberg, B., Rood, R., Hamilton, H. & Champion, F. Development of a performance criterion for enroute air traffic control personnel research through air traffic control simulation: Experiment I--parallel form development (FAA-RD-75-186, Interim Report). Atlantic City, NJ: Federal Aviation Administration, National Aviation Facilities Experimental Center, February 1976. (AD-A023 411).

Burgin, G.H. & Fogel, L.J. Air-to-air combat tactics synthesis and analysis program based on an adaptive maneuvering logic. Journal of Cybernetics, 1972, 2 (4), 60-68.

Burington, R.S. Concerning the reliability and effectiveness of weapon systems and their measurement (R-14-6). Washington, DC: Bureau of Naval Weapons, Office of Chief Mathematician, April 1961.

Campbell, S.C., Feddern, J. & Graham, G. A-6E Systems approach to training: Phase I (NAVTRAEQUIPC-75-C-0099-1). Bethpage, NY: Grumman Aerospace Corporation, February 1977. (AD-A037 468).

Chaikin, G. Stinger human factors engineering final report (Technical Memorandum 25-76) Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory, July 1976. (AD-B014 866).

Chaney, F.B., & Thresh, J.L. Diagnosis and correction of quality problems: A human factors approach. Paper presented at the ASQC Seminar on Product Quality Audit, Milwaukee, WI, April 19-20, 1968. (AD-855 919).

Chapanis, A. Relevance of physiological and psychological criteria to man-machine systems: The present state of the art. Ergonomics, 1970, 13, 337-346. (AD-751 344).

Chasteen, C.L. IOT&E of an AWADS radar imagery recorder. (MAC Project 74C-110U) Eglin AFB, FL: Military Airlift Command, Operating Location F, May 1975. (AD-B003 562).

Chop, A. Capability measures for system effectiveness (RADC-TR-72-26, Final Technical Report). Sunnyvale, CA: Lockheed Missiles and Space Company, February, 1972. (AD-892 863).

Churchman, C.W. Systems analysis and organization theory: A critique (Internal Working Paper No. 3). Berkeley, CA: University of California, Space Sciences Laboratory, June 1971.

City of Reading. Systems analysis completion report, Vol. 2: Systems analysis methodology (USAC-RPAO-005). Reading, PA: City of Reading, September 1971. (PB-208 500-2).

Clovis, E.R. & Muller, T.H. Development of procedures for evaluating unit performance (TRA-75/009, Final Report, Vol. 1). Monterey, CA: Litton Mellonics Defense Sciences Laboratories, March 1975.

Coburn, R. Human engineering guide to ship system development (NELC Technical Document 278). San Diego, CA: Naval Electronics Laboratory Center, October 1973. (AD-772 535).

Cogan, E.A. & Lyons, J.D. If it exists, it can be measured—but how? In E.A. Cogan Frameworks for measurement and quality control (HumRRO-PP-16-72). Papers presented at New York University First National Annual Training in Business and Industry Conference, New York City, March 1972. (AD-748 081)

Companion, M.A. & Corso, M. Task taxonomy: Two ignored issues. In A.S. Neal & R.F. Palasek (Eds.), Proceedings of the Human Factors Society 21st Annual Meeting, San Francisco, October 17-20, 1977.

Companion, M. & Teichner, W.H. Application of task theory to task analysis Evaluation of validity and reliability using simple tasks (AFOSR-TR-77-1008). Las Cruces, NM: New Mexico State University, Human Performance Laboratory, January 1977. (AD-A043 243)

Connelly, E.M., Bourne, F.J., Loental, D.G. & Knoop, P.A. Computer-aided techniques for providing operator performance measures (AFHRL-TR-74-87). McLean, VA: Quest Research Corporation, December 1974. (AD-A014 330).

Connelly, E.M., Bourne, F.J., Loental, D.G., Migliaccio, J.S., Burchick, D.A. & Knoop, P.A. Candidate T-37 pilot performance measures for five contact maneuvers (AFHRL-TR-74-88). McLean, VA: Quest Research Corporation, December 1974. (AD-A014 331).

Connelly, E.M., Schuler, A.R. & Knoop, P.A. Study of adaptive mathematical models for deriving automated pilot performance measurement techniques (AFHRL-TR-69-7, Vol. I: Model Development). Falls Church, VA: Melpar, October 1969. (AD-704 597).

Connelly, E.M. & Sloan, N.A. Manned system design using operator measures and criteria (OTR-62-76-1). Vienna, VA: Omnemii, Inc., October 1976. (AD-A032 687).

Connelly, E.M. & Sloan, N.A. Manned system design using operator measures and criteria. In A.S. Neal & R.F. Palasek (Eds.), Proceedings of the Human Factors Society 21st Annual Meeting. San Francisco, October 17-20, 1977.

Connelly, E.M., Zeskind, R.M. and Chubb, G.P. Development of a continuous performance measure for manual control. (AMRL-TR-76-24) Vienna, VA: Omnemii, Inc., April 1977. (AD-A 041 676).

Connelly, M.N. & Willis, J.E. The development of a human effectiveness junction allocation methodology (HEFAM) (Research Memorandum SRM 70-11). San Diego, CA: Navy Personnel and Training Research Laboratory, October 1969. (AD-699 173).

Cunningham, J.B. A systems-resource approach for evaluating organizational effectiveness. Human Relations, 1978, 31, (7), 631-656.

Cunningham, R.P. Sheldon, M.S. & Zagorski, H.J. Project norm: Pilot study report (TM-2232/000/00). Santa Monica, CA: System Development Corporation, February 1965. (AD-458 341).

Defense Documentation Center. Measurement methodology (A report bibliography). Alexandria, VA: Defense Documentation Center, February 1977. (Search Control No. 052637).

Defense Documentation Center. Measurement theory (A report bibliography). Alexandria, VA: Defense Documentation Center, February 1977. (Search Control No. 052638).

Defense Documentation Center. Operational test and evaluation (A report bibliography). Alexandria, VA: Defense Documentation Center, February 1977. (Search Control No. 052632).

Defense Documentation Center. Performance measurement (A report bibliography). Alexandria, VA: Defense Documentation Center, February 1977. (Search Control No. 052631).

Defense Documentation Center. *Performance standards measurement* (A report bibliography). Alexandria, VA: Defense Documentation Center, February 1977. (Search Control No. 052629).

Defense Documentation Center. *System human operators* (A report bibliography). Alexandria, VA: Defense Documentation Center, February 1977. (Search Control No. 052635).

Defense Documentation Center. *Systems and methodology* (A report bibliography). Alexandria, VA: Defense Documentation Center, February 1977. (Search Control No. 052639).

Dieterly, D.L. *The evaluation of training with specific emphasis on criteria* (AU-AFIT-SL-9-73). Wright-Patterson AFB, OH: Air Force Institute of Technology, School of Systems and Logistics, October 1973. (AD-771 009).

DiGialleonardo, F.R.& Barefoot, D.B. *An approach for measuring benefit and cost in management and information systems* (TR 75-21). San Diego, CA: Navy Personnel Research and Development Center, October 1974. (AD-A014 209).

DiGialleonardo, F.R., Barefoot, D.B. & Blanco, T.A. *Technique for interactive systems analysis (TISA).* (NPRDC-TR-75-22). San Diego, CA: Navy Personnel Research and Development Center, 1974. (AD-A013 223).

Duning, K.E., Hickok, C.W., Emerson, K.C. & Clement, W.F. *Control-display testing requirements study* (AFFDL-TR-72-122). Cedar Rapids, IA: Collins Radio Company, December 1972. (AD-759 539).

Dunlap, J.W. & Affinito, F.J. *Development of methodology for measuring effects of personal clothing and equipment on combat effectiveness of individual soldiers* (Final Report of Phase 3). Darien, CT: Dunlap and Associates, Inc., December 1967. (AD-836 904).

Dunlap and Associates, Inc. *Performance measures for human factors engineering evaluation of EARL equipment* (SSD-332-618). Darien, CT: Dunlap and Associates, Inc., June 1966. (AD-653 627).

Egbert, J.J. & Rau, J.G. *Techniques for bounding and estimating measures of effectiveness.* Newport Beach, CA: Ultrasystems, Inc., July 1973. (AD-772 550).

Ellis, R.H. *Task analysis reduction technique (TART) for the quantification of human performance* (Research Memorandum SRM 71-7). San Diego, CA: Naval Personnel and Training Research Laboratory, September 1970. (AD-711 807).

Engel, J.D. *An approach to standardizing human performance assessment* (Professional Paper 26-70). Presentation at the Planning Conference of Standardization of Tasks and Measures for Human Factors Research. Texas Technological University, Lubbock, TX: March 1970. (AD-717 258).

Eppler, W.G. Analytical design of manned control systems (SUDAAR No. 280). Stanford, CA: Stanford University, Center for Systems Research, May 1966.

Erickson, R.A. Field evaluation of a visual detection model. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Farina, A.J. & Wheaton, G.R. Development of a taxonomy of human performance: The task characteristics approach to performance prediction (Technical Report 7). Pittsburgh, PA: American Institutes for Research, February 1971. (AD-736 191).

Featherstone, C.L. & Scaglione, R.J. A feasibility study for determining a small arms measure of effectiveness for handling characteristics (Master's Thesis). Monterey, CA: Naval Postgraduate School, September 1975. (AD-B008 586L).

Fineberg, M.L., Meister, D., and Farrell, J.P. An assessment of the navigation performance of army aviation under nap-of-the-earth conditions (Research Report 1195) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A060 563).

Finley, D.L. & Muckler, F.A. Human factors research and the development of a manned systems applications science: The system sampling problem and a solution. Northridge, CA: Manned Systems Sciences, Inc., July 1976.

Finley, D.L., Muckler, F.A., Gainer, C.A. & Obermayer, R.W. An analysis and evaluation methodology for command and control: Final technical report. Northridge, CA: Manned Systems Sciences, Inc., November 1975. (AD-A023 871).

Finley, D.L., Obermayer, R.W., Bertone, C.M., Meister, D. & Muckler, F.A. Human performance prediction in man-machine systems —Volume I—A technical review. (NASA CR-1614) Canoga Park, CA: The Bunker-Ramo Corporation, August 1970.

Fischl, M.A., Siegel, A.I., & Wolf, J.J. Application of a multiple task interactive model: Simulation of human performance in sonar maintenance. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Foley, J.P. Criterion referenced measures of technical proficiency in maintenence activities (Final Report). Wright-Patterson AFB, OH: Air Forces Human Resources Laboratory, Advanced Systems Division, October 1975. (AD-A016 420).

Foley, J.P., Jr. Performance measurement of maintenance. Wright-Patterson Air Force Base, OH: Advanced Systems Division, Air Force Human Resources Laboratory, 1977. (AD-A053 475)

Ford, J.P., Harris, J.H. & Rondiac, P.F. Performance measures for AIT armor crewman (HumRRO-CR-D2-74-2). Alexandria, VA: Human Resources Research Organization, April 1974. (AD-A019 375).

Geddie, J.C. Profiling the characteristics of the developmental test participant (Technical Memorandum 31-76). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory, October 1976. (AD-A031 563).

Geer, C.W. Navy manager's guide for the test and evaluation sections of MIL-H-46855 (D194-10006-2). Seattle, Washington, Boeing Aerospace Company, Logistics Support and Services, June 30, 1977. (AD-A 045 098).

Geer, C.W. User's guide for the test and evaluation sections of MIL-H-46855 (D194-10006-1). Seattle, Washington: Boeing Aerospace Company, Logistics Support and Services, June 30, 1977. (AD-A045 097).

George, C.E. & Dudek, R.A. Performance, recovery and man-machine effectiveness: Final report on a basic research program under project THEMIS (Technical Memorandum 9-74). Lubbock, TX: Texas Technical University, April 1974. (AD-777 797).

Gephart, L.S. & Balachandran, V. A stochastic system effectiveness simulation model (Technical Report UDRI-TR-70-12). Dayton, OH: University of Dayton-Research Institute, June 1969. (AD-865 449).

Gex, R.C. Personnel subsystem testing and evaluation for missiles and space systems: An annotated bibliography (SB-61-21). Sunnyvale, CA: Lockheed Aircraft Corporation, Missiles and Space Division, April 1961.

Goldbeck, R.A. Wright, K.A. & Fowler, R.L. Operator performance and panel layout for discontinuous tasks (AMRL-TR-70-137). Palo Alto, CA: Philco-Ford Corporation., WLD Division - Human Engineering Section, March 1971. (AD-727 791).

Greening, C.P. Unfinished business in the utility of visual detection models. G.W. Levy (Ed.) Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Gruenberger, F.J. A measure for crackpots. Santa Monica, CA: The RAND Corporation, December 1962. (AD-293 264).

Grunzke, P.M. Evaluation of the automated adaptive flight training system's air-to-air intercept performance measurement. (AFHRL-TR-78-23) Williams Air Force Base, AZ: Flying Training Division, July 1978. (AD-A060 320)

Gustafson, H.W. The measurement of proficiency of air force maintenance personnel (ATPTRC SM 7-1957) Denver, CO: Air Force Personnel and Training Research Center, Air Research and Development Command, Lowry Air Force Base, 29 July 1967. (AD-841 711).

Haight, F.A. Indirect methods for measuring exposure factors as related to the incidence of motor vehicle traffic accidents (DOT-HS-800-601). State College, PA: Pennsylvania State University, September 1971. (PB-205 031)

Hakanson, N.H. An adaptive method of test selection in system development (RM-5238-PR). Santa Monica, CA: The RAND Corporation, April 1967. (AD-653 942).

Hall, D.H. Confidence limits for measures of effectiveness in weapon system analysis (NWC TP 5437). China Lake, CA: Naval Weapons Center, May 1973. (AD-762 401).

Hammell, T.J., Gasteyer, C.E. & Pesch, A.J. Advanced officer tactics training device needs and performance measurement technique (NAVTRAEQUIPCEN72-C-0053-1, Vol. 1). Groton, CT: General Dynamics/Electric Boat Division, November 1973. (AD-922 929L).

Hanifan, D.T. & Knowles, W.B. Human performance in system effectiveness modeling: Issues, approaches and critique. Santa Monica, CA: Dunlap and Associates, Inc., December 1968.

Hansen, D.N., Johnson, B.F., Fagan, R.L., Tam, P., and Dick, W. Computer-based adaptive testing models for the Air Force technical training environment Phase I: Development of a computerized measurement system for air force technical training (AFHRL-TR-74-48) Tallahassee, FL: Florida State University, July 1974. (AD-785 142).

Hansen, D.N., Ross, S. & Harris, D.A. Flexilevel adaptive testing paradigm: Validation in technical training (AFHRL-TR-77-35-1-). Memphis, TN: Memphis State University, July 1977.

Harrison, W.L. A theoretical basis for the concept of effectiveness (Master's Thesis). Monterey, CA: Naval Post Graduate School, October 1966. (AD-807 386).

Harry, H.P. RANN utilization experience. Case study number 10. Effectiveness measurement methods. (NSF-RA-G-75-038). Washington. D.C.: Urban Institute, 1975. (PB-247 254).

Helm, W.R. Human engineering design deficiencies: A comparative analysis of the P-3 and S-3 aircraft. (NATC-SY-41R-76). Patuxent River, MD: Naval Air Test Center, March 1976. (AD-B010 940).

Helm, W.R. Human factors test and evaluation, functional description inventory as a test and evaluation tool, development and initial validation study (NATC-SY-77R-75 Vol. 1). Patuxent River, MD: Naval Air Test Center, September 1975. (AD-B007 463L).

Henderson, R.L. & Burg, A. The role of vision and audition in truck and bus driving (Final Report). Santa Monica, CA: System Development Corporation, December 1973. (PB-230 776).

Hicks, J.A. A methodology for conducting human factors evaluations of vehicles in operational field tests (Draft). Fort Hood, TX: U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Hood Field Unit, June 1977.

Hicks, J.A. A methodology for conducting human factors evaluations of vehicles in operational field tests. In A.S. Neal & R.F. Palasek (Eds.), Proceedings of the Human Factors Society 21st Annual Meeting. San Francisco, October 17-20, 1977.

Highsmith, R.C. Proposed measure of effectiveness for human resource availability periods and their impact upon unit operational readiness. Monterey, CA: Naval Postgraduate School, December 1976. (AD-A036 025).

Hill, J.W. & Eddowes, E.E. Further development of automated GAT-1 performance measures (AFHRL-TR-73-72). Menlo Park, CA: Stanford Research Institute, May 1974. (AD-783 240).

Holshouser, E.L. Guide to human factors engineering general purpose test planning (GPTP) (TP-77-14) Point Mugu, CA: Pacific Missile Test Center, June 1977. (AD-BO22 013L).

House, E.R. Assumptions underlying evaluation models. Educational Researcher, March 1978, 7 (3), 4-12.

Hyatt, C.J. & DeBerg, O.H. A scoring system for the quantitative evaluation of pilot performance during microwave landing system (MLS) approaches (ASD-TR-17). Wright-Patterson AFB, OH: Aeronautical Systems Division (AFSC), August 1975. (AD-A025 782).

Hunter, B. Report on the final acceptance test for CTS (Computerized Training System). Alexandria, VA: Human Resources Research Organization, August 1976. (AD-A033 400),

Hutchins, C.W. A computer aided function allocation and evaluation system (CAFES). Proceedings of the Human Factors Society Annual Meeting, Huntsville, AL, October 15-17, 1974.

Irish, P.A., Grunzke, P.M., Gray, T.H. & Waters, B.K. The effects of system and environmental factors upon experienced pilot performance in the advanced simulator for pilot training (AFHRL-TR-77-13). Williams AFB, AZ: Air Force Human Resources Laboratory, Flying Training Division, April 1977.

Jahns, D.W. A concept of operator workload in manual vehicle operations. Meckenheim, BRD: Forschungsinstitut fur Anthropotechnik, December 1973.

Jahns, D.W. & Katz, R. Preliminary concepts for computer aided function-allocation evaluation system (CAFES) (NADC-72106-CS). Seattle, WA: The Boeing Company, Aerospace Group, May 1972. (AD-902 670L).

James, L.R. Criterion models and construct validity for criteria (72-32). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, August 1972. (AD-770 412).

Jaschen, D.G. Mechanized infantry combat vehicle, XM723, system operational climatic test. (TDP-OT-67A) Falls Church, VA: U.S. Army Operational Test and Evaluation Agency, Test Design Division, 1975. (AD-B008 212L).

Kagerer, R.L. & Weiss, E.C. Development of a checklist and guidebook for human factors evaluation of general equipment (Final Report). Alexandria, VA: The Matrix Corporation, January 1968. (AD-827 808).

Kaplan, J.D., Crooks, W.H., Boylan, R.J. & Sanders, M.S. Human resources test and evaluation system (HRTES) (PDR-1057-78-111). Woodland Hills, CA: Perceptronics, November 1978.

Karaush, A.D. Benchmark analysis of time-sharing systems (SP-3347). Santa Monica, CA: System Development Corporation, June 1969. (AD-689 781).

Karush, A.D. Two approaches for measuring the performance of time-sharing systems (SP-3364). Santa Monica, CA: System Development Corporation, May 1969. (AD-691 366).

Kelley, C.R. Human operator models for manual control. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Kerner, H. and Beyerle, W. A PMS level language for performance evaluation modelling (V-PMS). Vienna: Institut fuer Digitale Anlagen Technical University.

Kiraly, R.J., Babinsky, A.D. & Powell, J.D. Aircrew oxygen system development: Man-in-the loop test report (NASA CR-73395). Cleveland, OH: TRW Inc., July 1970.

Klein, R. Integrated operational test and analysis procedures for small arms weapon systems evaluation. Manuscript in preparation (undated).

Klein, R.D. & Thomas, C.B. The development of combat related measures for small arms evaluation. Ft. Benning, GA: U.S. Army Infantry Board, June 1969. (AD-713 552).

Kleinman, D.L. & Baron, S. Manned vehicle systems analysis by means of modern control theory (NASA CR-1753). Cambridge, MA: Bolt, Beranck and Newman, Inc., June 1971.

Knoop, P.A. Survey of human operator modeling techniques for measurement applications (AFHRL-TR-78-35). Wright-Patterson AFB, OH: Air Force Human Resources Laboratory, Advanced Systems Division, July 1978.

Knowles, W.B., Burger, W.J., Mitchell, M.B., Hanifan, D.T. & Wulfeck, J.W. Models, measures, and judgments in system design. Human Factors, 1969, 11, 577-590.

Kreifeldt, J. Analysis of predictor model. Unpublished Preliminary Data, Massachusetts Institute of Technology, 1964.

Kribs, H.D. & Thurmond, P. Computerized collective training for teams: Team job/task and training analysis (Second Interim Report). San Diego, CA: Sensors, Data, Decisions, Inc. March 1977.

Kvalseth, T.O. A decision-theoretic model of the sampling behavior of the human process monitor. IEEE Transactions on Systems, Man, and Cybernetics, November 1977, SMC-7, (11).

Labor/Management Task Force on Rail Transportation. Terminal performance measurement systems. (DOT-FR-4-3003). Washington, D.C., May 1975. (PB-252 225).

Larson, O.A., Sander, S.I. & Steinemann, J.H. Survey of unit performance effectiveness measures (TR74-11). San Diego, CA: Navy Personnel Research and Development Center, January 1974. (AD-774 919).

Lawrence, D.H. The evaluation of training and transfer programs in terms of efficiency measures. The Journal of Psychology, 1954, 36, 367-382. (AD-87 172).

Levy, G.W. Criteria for selection and application of models. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Lindsey, M.B. A human performance investigation in a NAVORD weapon system—Part 2 (Report 27-349). Annapolis, MD: Naval Ship Research and Development Center, Propulsion and Auxiliary Systems Department, March 1974. (AD-918 202L).

Lomov, B.F. [Systemic approach to psychology.] Voprosy Psikhologii. 1975 2, 31-45. (Arlington, VA: Joint Publications Research Service, 1975, JPRS 65806.)

LTV Aerospace Corporation. Shuttle active thermal control system. Vol. 3: Modular radiator system test date correlation with thermal model (Report No. T169-28, NASA CR-M0269). Dallas, TX: LTV Aerospace Corporation, Vought Systems Division, December 1973.

Lyons, J.D. Measuring effectiveness: Quality control of training. In E.A. Cogan & J.D. Lyons, Framework for measurement and quality control. (HumRRO-PP-16-72). Papers presented at New York University First National Annual Training in Business and Industry Conference, New York City, March 1972. (AD-748 081).

Markel, G.A. Toward a general methodology for systems evaluation (352-R-13). State College, PA: HRB-Singer, Inc., July 1965. (AD-619 373).

Mason, A.K. & Rigney, J.W. Toward a general characterization of electronic troubleshooting. In R.E. Blanchard & D.H. Harris (Eds.), Man-machine effectiveness analysis. Papers presented at The Human Factors Symposium, University of California at Los Angeles, June 15, 1967. (AD-735 718).

Matheny, W.G., Patterson, G.W. & Evans, G.I. Human factors in field testing (LS-FR-71-1). Hurst, TX: Life Sciences, Inc., December 1971. (AD-741 216).

Mathews, E.P. A visual target acquisition model. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

McCalpin, J.P. Development of a methodology for obtaining human performance reliability data for a machine gun system (Technical Memorandum 6-74). Baltimore, MD: AAI Corporation, February 1974. (AD-777 020).

McDonnell Douglas Astronautics Company, Eastern Division. Optimized cost/performance design methodology, Vol. 3—Concept analysis and model development, Book 1—Concept analysis (MDC-E0005). September 1969.

McDonnell Douglas Astronautics Company, Eastern Division. Optimized cost/performance design methodology, Vol. 3—Concept analysis and model development, Book 2—Model Formulation (MDC-E0005). September 1969.

McGrath, J.E., Nordlie, P.G. & Vaughan, W.S. A systematic framework for comparison of system research methods (HRS-TN-59/7-SM). Arlington, VA: Human Sciences Research, Inc., November 1959. (AD-229 923).

McKendry, J.M., and Harrison, P.C. Assessing human factors requirements in the test and evaluation stage of systems development, Volume II. (ND64-68) State College, PA: HRB-Singer, Inc., June 1964. (AD-603 304).

McMaster. Soviet night operations. U.S. Army Aviation Digest, January 1976, pp. 3; 14-17.

McRuer, D., Jex, H.R., Clement, W.F. & Graham, D. A systems analysis theory for displays in manual control (STI Technical Report 163-1). Hawthorne, CA: Systems Technology, Inc., June 1968. (AD-675 983).

Meister, D.A. A pragmatic approach to the prediction of operational performance. In R.E. Blanchard & D.H. Harris (Eds.), Man-machine effectiveness analysis. Papers presented at The Human Factors Symposium, University of California at Los Angeles, June 15, 1967. (AD-735 718).

Meister, D. Assumptions underlying test and evaluation strategies. San Diego, CA: Navy Personnel Research and Development Center. (Speech).

Meister, D. Assumptions underlying the human reliability model. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Meister, D. A systematic aproach to human factors measurement. San Diego, CA: Navy Personnel Research and Development Center, October 1978.

Meister, D. Human factors in operational system testing: A manual of procedures (NPR DC SR 78-8). San Diego, CA: Navy Personnel Research and Development Center, April 1978.

Meister, D. & Rabideau, G.F. Human factors evaluation in system development. New York: John Wiley & Sons, 1965.

Melching, W.H. A concept of the role of man in automated systems (Professional Paper 14-68). Presented at the Southwestern Psychological Association Annual Meeting, New Orleans, LA, April 1968. (AD-671 128).

Meyer, R.P., Laveson, J.I. & Pape, G.L. Development and application of a task taxonomy for tactical flying (AFHRL-TR-42, Vol. 1). St. Louis, MO: Design Plus, September 1978. (AD-061 387).

Meyer, R.P., Laveson, J.I. & Pape, G.L. Development and application of a task taxonomy for tactical flying (AFHRL-TR-78-42, Vol. 2). St. Louis, MO: Design Plus, September 1978. (AD-061 388).

Meyer, R.P., Laveson, J.I. & Pape, G.L. Development and application of a task taxonomy for tactical flying (AFHRL-TR-42, Vol. 3). St. Louis, MO: Design Plus, September 1978. (AD-A061 478).

Miles, J.L. Some recent trends in human factors testing (Technical Memorandum 22-76). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory, June 1976. (AD-B012 110).

Miller, J.G. Living Systems. New York: McGraw-Hill, 1978.

Miller, J.R. Assessing alternative transportation systems (RM-5865-DOT). Santa Monica, CA: The RAND Corporation, April 1969. (PB-185 167).

Mills, R.G. & Hatfield, S.A. Sequential task performance: Task module relationships, reliabilities, and times (AMRL-TR-72-56). Wright-Patterson AFB, OH: Aerospace Medical Research Laboratory, 1974. (AD-787 322).

Mitchell, M.B. & Blanchard, R.E. The allocation of system effectiveness requirements for man-machine effectiveness analysis. In R.E. Blanchard & D.H. Harris, (Eds.), Man-machine effectiveness analysis. Papers presented at the Human Factors Symposium, University of California at Los Angeles, June 15, 1967. (AD-735 718).

Mitchell, M.B., Smith, R.L. & Blanchard, R.E. Test application of TEPPS on a Navy CIC subsystem. Santa Monica, CA: Dunlap and Associates, Inc., August 1967. (AD-821 573).

Montgomery, D.C., Callahan, L.G. & Wadsworth, H.M. Application of decision/risk analysis in operational tests and evaluation. Atlanta, GA: Georgia Institute of Technology, The School of Industrial and Systems Engineering, September 1975. (AD-A024 205).

Mumford, J. & Smith, J.P. The development of performance criteria for turret mechanics (HumRRO Research Memorandum). Alexandria, VA: Human Resources Research Organization, July 1961. (AD-477 647).

Narva, M.A. Development of a systematic methodology for the application of judgmental data to the assessment of training device concepts. Paper presented at the U.S. Army Science Conference, West Point, NY, June 20-22, 1978.

Nelson, R.T. A research methodology for studying complex service systems (Working Paper No. 139). Los Angeles, CA: University of California, Western Management Science Institute, July 1968. (AD-673 233).

Obermayer, R.W. & Vreuls, D. Combat-ready crew performance measurement system: Phase I measurement requirements (AFHRL-TR-108-II). Northridge, CA: Manned Systems Sciences, Inc., December 1974. (AD-B005 518L).

Obermayer, R.W. & Vreuls, D. Combat-ready crew performance measurement system: Phase II measurement system requirement. Northridge, CA: Manned Systems Sciences, Inc. December 1974. (AD-B005 519).

Obermayer, R.W. & Vreuls, D. Combat-ready crew performance measurement system: Phase IIIA crew performance measurement (AFHRL-TR-74-108-IV). Northridge, CA: Manned Systems Sciences, Inc., December 1974. (AD-B005 520L).

Obermayer, R.W., Vreuls, D. & Conway, E.J. Combat-ready crew performance measurement system: Phase IIIC design studies (AFHRL-TR-74-108-VI). Northridge, CA: Manned Systems Sciences, Inc., December 1974. (AD-B005 521L).

Obermayer, R.W., Vreuls, D., Muckler, F.A. & Conway, E.J. Combat-ready crew performance measurement system: Phase IIID specifications and implementation plan. Northridge, CA: Manned Systems Sciences, Inc., December 1974. (AD-B005 522L).

O'Connor, M.F. & Buede, D.M. The application of decision analytic techniques to the test and evaluation phase of the acquisition of a major air system (TR-77-3). McLean, VA: Decisions and Designs, Inc., April 1977. (AD-A040 691).

Performance Measurement Associates, Inc. Team performance measures for computerized systems (Unsolicited technical proposal PMA-022). Vienna, VA: PMA, Inc., July 1978.

Pew, R.W., Baron, S., Feehrer, C.E. & Miller, D.C. Critical review and analysis of performance models applicable to man-machine systems evaluation (BBN Report No. 3446). Cambridge, MA: Bolt Beranek and Newman, Inc., March 1977. (AD-A038 597).

Phatak, A.V. Improvement in weapon system effectiveness by application of identification methods for determining human operator performance decrements under stress conditions (AMRL-TR-73-38). Palo Alto, CA: Systems Control, Inc., December 1973. (AD-773 856).

Phatak, A., Weinert, H. & Segall, I. Identification of the optional control model for the human operator (AMRL-TR-74-79). Palo Alto, CA: Systems Control, Inc., May 1975. (AD-A009 956).

Philco-Ford Corporation. Human engineering design check list (WDL-TR 1968A). Palo Alto, CA: Philco-Ford Corporation, May 1964. (AD-829 426).

Polak, W.T., Robertson, J.C. & Yuan, L.S. Systems effectiveness analyzer (AMRL-TR-73-113). Sudbury, MA: Raytheon Company, Simulation and Surveillance Systems, February 1974. (AD-781 124).

Popham, W.J. & Husek, T.R. Implications of criterion-referenced measurement. Spring 1969.

Pritsker, A.A., Wortman, D.B., Seum, C.S., Chubb, G.P. & Seifert, D.J. SAINT: Vol. 1. Systems analysis of integrated networks of tasks (AMRL-TR-73-126). West Lafayette, IN: Purdue University, School of Industrial Engineering, April 1974. (AD-A014 843).

Quinn, J.L. Research and development appraisal and evaluation (SLTR5-70). Wright-Patterson AFB, OH: Air University, School of Systems and Logistics, November 1970. (AD-876 006).

Ramsay, J.O. Some statistical considerations in multidimensional scaling (RB-66-26). Princeton, NJ: Educational Testing Service, June 1966. (AD-489 591).

Rankin, G.L. An application of fault free analysis to operational testing (Master's Thesis). Georgia Institute of Technology, 1975.

Rankine, R.R. The effects of aircraft dynamics and pilot performance on tactical weapon delivery accuracy (UCLA-ENG-7085). Los Angeles, CA: University of California, School of Engineering and Applied Science, November 1970. (AD-728 324).

Rasch, W.H.. Jr. Guidelines for making tradeoffs: The special role of technical performance measurement. Ft. Belvoir, VA: Defense Systems Management College, 1973. (AD-A045 256).

Rau, J.G. Measures of effectiveness handbook. Irvine, CA: Ultra Systems, Inc., August 1974. (AD-A021 461).

Ray, E.G., Hiltz, H.A., and Spurway, D.T. Test design plan for defense satellite communications system III (DSCS III) initial operational test and evaluation (IOT&E). U.S. Army Operational Test and Evaluation Agency, Test Design Division, 1979. (AD-B041 308L).

Repperger, D.W., Smiles, K.A., Neff, J.A. & Summers, W.C. A feature selection approach in evaluation of parameter changes on the human operator under thermal stress. Ergonomics, 1978, 21, (1), 35-48.

Rhoads, D.W. Inflight evaluation of four cockpit controller configurations in a variable stability airplane (AFFDL-TR-70-95). Buffalo, NY: Cornell Aeronautical Laboratory, Inc., September 1970. (AD-876 589).

Rigby, L.V. The Sandia Human Error Rate Bank (SHERB). In R.E. Blanchard & D.H. Harris (Eds.), Man-machine effectiveness analysis. Papers presented at the Human Factors Symposium, University of California at Los Angeles, June 15, 1967. (AD-735 718).

Rouse, W.B. Estimation and control theory: Application to modeling human behavior. Human Factors, 1977, 19, (4), 315-329.

Sauer, D.W., Campbell, W.B., Potter, N.R., and Askren, W.B. Human resource factors and performance relationships in nuclear missile handling tasks. (AFHRL-TR-76-85, AFWL-TR-76-301) Dayton, OH: Systems Research Laboratories, Inc., 1977. (AD-A042 604).

Schaeffer, K.H., Fink, J.B., Rappaport, M., Wainstein, L. & Erickson, C.J. The knowledgeable analyst: An approach to structuring man-machine systems (SRI No. IMU-3546). Menlo Park, CA: Stanford Research Institute, February 1963. (AD-297 432).

Self, H.C. Performance measures, observer selection, and reconnaissance/strike effectiveness (AMRL-TR-72-86). Wright-Patterson AFB, OH: Aerospace Medical Research Laboratory, November 1972. (AD-770 647).

Sheldon, M.S. & Zagorski, H.J. Man-machine system evaluation—The normative operations reporting method. In R.E. Blanchard & D.H. Harris (Eds.), Man-machine effectiveness analysis. Papers presented at The Human Factors Symposium, University of California at Los Angeles, June 15, 1967. (AD-735 718).

Sidoruk, W. Methodology for evaluation of image enhancement techniques. Human Factors, 1977, 19, (3), 263-272.

Siegel, A.I. Human performance reliability—It's measurement and impact on system reliability. ASME Publication, 78-DE-17, New York: The American Society of Mechanical Engineers, 1978.

Siegel, A.I. & Federman, P.J. Development of performance evaluative measures: Investigation into and application of a fleet post-training performance evaluation system (Report No. 7071-2). Wayne, PA: Applied Psychological Services, Inc., September 1970. (AD-713 192).

Siegel, A.I., Pfeiffer, M.G. & Worms, T.M. Development and evaluation of a content analytic approach in Army field system data organization (ARI Technical Report No. TR-77-A11). Wayne, PA: Applied Psychological Services, Inc., August 1977. (AD-A042 075).

Siegel, A.I. Platzer, H.L. & Lauterman, R.S. Techniques for evaluating operator loading in man-machine systems. Wayne, PA: Applied Psychological Services. (AD-810 787).

Siegel, A.I., Schultz, D.G. & Federman, P. Post-training performance criterion development and application: A matrix method for the evaluation of training. Wayne, PA: Applied Psychological Services, Inc., January 1961. (AD-251 082).

Siegel, A.I. & Wolf, J.J. Computer simulation: Savior, sanctuary, or silliness? In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Siegel, A.I. Wolf, J.J. & Lautman, M.R. A model for predicting integrated man-machine system reliability: Model logic and description. Wayne, PA: Applied Psychological Services, Inc., November 1974. (AD-A009 814).

Simon, C.W. Advanced methodologies study program (Final Report). Culver City, CA: Hughes Aircraft Company, June 1975. (AD-A021 099).

Simon, D.E. et al. Stanford procedures for Air Force operational test and evaluation (RADC-TR-74-270, Final Report, Vol. 1). Albaquerque, NM: Braddock, Dunn & McDonald, Inc., October 1974. (AD-B000 365).

Sjogren, D.D. Measurement techniques in evaluation. Review of Educational Research, 40, 301-320.

Smith, R.L., Westland, R.A. & Blanchard, R.E. Technique for establishing personnel performance standards—TEPPS (PTB-70-5, Vol. 1). Santa Monica, CA: Integrated Sciences Corporation, December 1969. (AD-704 103).

Smith, R.L., Westland, R.A. & Blanchard, R.E. Technique for establishing personnel performance standards--TEPPS (PTB-70-5, Vol. 2). Santa Monica, CA: Integrated Sciences Corporation, December 1969. (AD-704 104).

Spencer, G. Man-machine simulation—the PIMO application. In R.E. Blanchard & D.H. Harris (Eds.), Man-machine effectiveness analysis. Papers presented at The Human Factors Symposium, University of California at Los Angeles, June 15, 1967. (AD-735 718).

Spyker, D.A., Stackhouse, S.P., Khalafalla, A.S. & McLane, R.C. Development of techniques for measuring pilot workload (NASA CR-1888). Roseville, MN: Honeywell Inc., November 1971.

Swink, J.R., Butler, E.A., Lankford, H.E., Miller, R.M., Watkins, H. & Waag, W.L. Definition of requirements for a performance measurement system for C-5 aircrew members (AFHRL-TR-78-54). San Diego, CA: Logicon Inc., October 1978.

Taylor, E.N. & Tillman, B. Human factors engineering study of two ball port designs for IFV. Unpublished manuscript, October 1977.

The Bunker-Ramo Corporation. Proposal for a study of human factors field performance measurement methodology (Proposal No. 5655-022-5U1). Canoga Park, CA: The Bunker-Ramo Corporation, Defense Systems Division, June 1965.

Thurmond, P. Development of analysis, design and development technique for team ISD (Unsolicited proposal). San Diego, CA: PATH Research. (Undated).

Tien, J.M. Toward a systematic approach to program evaluation design. IEEE Transactions on Systems, Man, and Cybernetics, September 1979, SMC-9, (9), 494-515.

Timson, S.F. Measurement of technical performance in weapon system development programs: A subjective probability approach (RM-5207-ARPA). Santa Monica, CA: The RAND Corporation, December 1968. (AD-681 771).

Topmiller, D.A. Mathematical models of human performance in man-machine systems (AMRL-TR-68-22). Wright-Patterson AFB, OH: Aerospace Medical Research Laboratories, May 1968. (AD-673 348).

Topmiller, D.A. The role of applied man-machine models. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Tranby, E.D. Advanced surface ship weapon systems test and evaluation (NSWSES-TP-60). Port Hueneme, CA: Naval Ship Weapon Systems Engineering Station, January 1976. (AD-390 826).

Turner, C.R. & Bard, J.F. Tactical AWACS measures of effectiveness (ESD-TR-72-142). Bedford, MA: The MITRE Corporation, April 1972. (AD-742 233).

Uhlaner, J.E. Human performance, jobs, and systems psychology—the systems measurement bed (Technical Report S-2). Arlington, VA: Behavior and Systems Research Laboratory, October 1970. (AD-716 346).

Uhlaner, J.E. and Drucker, A.J. Military research on performance criteria: A change of emphasis. Human Factors, 1980, 22, (2), 131-139.

Ultrasystems, Inc. A compendium of measures of effectiveness used in Navy studies (Preliminary Version). Newport Beach, CA: Ultrasystems, Inc., 1971.

Ultrasystems, Inc. A study of measures of effectiveness used in Naval analysis studies (Vol. 1 Summary). Newport Beach, CA: Ultrasystems, Inc., October 1972. (AD-912 443).

Ultrasystems, Inc. A study of measures of effectiveness used in Naval analysis studies (Vol. 2 Study Review Summaries, Part 1). Newport Beach, CA: Ultrasystems, Inc., October 1972. (AD-912 444).

Ultrasystems, Inc. A study of measures of effectiveness used in Naval analysis studies (Vol. 3 Study Review Summaries, Part 2). Newport Beach, CA: Ultrasystems, Inc., October 1972. (AD-912 445).

Ultrasystems, Inc. A study of measures of effectiveness used in Naval analysis studies (Vol. 4 MOE Reviews). Newport Beach, CA: Ultrasystems, Inc., October 1972. (AD-912 446).

U.S. Army Armor School. Tank platoon: Organization for combat and techniques of movement (TC 17-15-3). Fort Knox, KY: U.S. Army Armor School, April 1975. (Pamphlet).

U.S. Army Combat Developments Command. Communications-electronics-75 (CE-75) phase 1-field Army (No. 6492, Vol. 9). Fort Monmouth, NJ: U.S. Army Combat Developments Command, Communications-Electronics Agency, September 1968. (AD-883 611).

U.S. Army Infantry Board. Infantry weapons test methodology study. Vol. 1: Small arms test methodology (USAIB-3319-F-Vol. 1). Fort Benning, GA: U.S. Army Infantry Board, November 1971. (AD-890 998L).

U.S. Army Infantry School. Individual-collective training plan (I-CTP) for XM723 mechanized infantry combat vehicle (rev.). U.S. Army Infantry School, June 1976.

U.S. Army Infantry School. Individual-collective training plan for XM723 mechanized infantry combat vehicle (rev.). U.S. Army Infantry School.

U.S. Army, Army Materiel Command. Engineering design handbook: Recoilless rifle weapon systems. Alexandria, VA. 15 January 1976. (AD-A023 513).

U.S. Army, Materiel Testing Directorate. U.S. Army test and evaluation command development Test II (EP)—System test operations procedure "test, measurement, and diagnostic equipment (system peculiar)." Aberdeen Proving Ground, MD, U.S. Army Human Engineering Laboratory, 7 May 1974. (AD-781

U.S. Army Operational Test and Evaluation Agency. Mechanized infantry combat vehicle (MICV), XM723, operational climatic test force, development test and experimentation (Draft). Falls Church, VA: U.S. Army Operational Test and Evaluation Agency, March 1976.

U.S. Army Test and Evaluation Command. Materiel test procedure 3-3-521. Fort Benning, GA: U.S. Army Infantry Board, May 1970. (AD-871 788).

U.S. Army Test and Evaluation Command. Materiel test procedure 6-2-502 (Electronics Proving Ground). Fort Benning, GA: U.S. Army Infantry Board, August 1969. (AD-720 976).

U.S. Army Test and Evaluation Command. Materiel test procedure 10-2-505 (General Equipment Test Activity). Fort Benning, GA: U.S. Army Infantry Board, September 1971. (AD-729 855).

U.S. Department of the Army. Army training and evaluation program for mechanized infantry battalion and combined arms task force (ARTEP 7-45). Washington, DC: Headquarters, Department of the Army, September 1975.

U.S. Department of the Army. Breathe easy, tanker or NBC collective protection for armored vehicles (No. TC 3-2). Washington, DC: Headquarters, Department of the Army, October 1977. (Pamphlet).

U.S. Department of the Army. Cavalry (Field Manual 17-95). Washington, DC: Headquarters, Department of the Army, July 1977.

U.S. Department of the Army. Machinegun 7.62-mm,M60 (Field Manual 26-67). Washington, DC: Headquarters, Department of the Army, October 1964.

U.S. Department of the Army. Skill qualification test notice for MOS 11B SQT 3 infantryman (No. 7-11B3-N). Washington, DC: Headquarters, Department of the Army, June 1976. (Pamphlet).

U.S. Department of the Army. Tank gunnery (Field Manual 17-12). Washington, DC: Headquarters, Department of the Army, March 1977.

U.S. Department of the Army. Tank gunnery devices (Field Manual 17-12-7). Washington, DC: Headquarters, Department of the Army, June 1977.

U.S. Department of the Army. Technique of fire of the rifle squad and tactical application (Field Manual 23-12). Washington, DC: Headquarters, Department of the Army, October 1967.

U.S. Department of the Army, Office of the Chief of Staff, Tank Forces Management Group. Tank weapon system management. Washington, DC.

Van Acker, A. & Wohl, J.G. Modeling the sonar operator's detection process: A progress report. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation. November 1968. (AD-697 939).

von Winterfeldt, D. An overview, integration, and evalution of utility theory for decision analysis (SSRI Research Report 75-9). Los Angeles, CA: University of Southern California, Special Science Research Institute, August 1975. (AD-A021 497).

Vreuls, D., Wooldridge, A.L., Obermayer, R.W., Johnson, R.M. & Norman, D.A. Development and evaluation of trainee performance measures in an automated instrument flight maneuvers trainer (NAVTRAEQUIPLEN 74-C-0063-1). Westlake Village, CA: Canyon Research Group, Inc., October 1975. (AD-A024 517).

Waag, W.L., Eddowes, E.E., Fuller, J.H. & Fuller, R.R. ASUPT automated objective performance measurement system (AFHRL-TR-75-3). Williams AFB, AZ: Air Force Human Resources Laboratory, Flying Training Division, March 1975. (AD-A014 799).

Weapon System Effectiveness Industry Advisory Committee. Final report of task group II. Vol. 1 Prediction-measurement: Summary, conclusions and recommendations (AFSC-TR-65-2). 1965. (AD-458 454).

Weapon System Effectiveness Industry Advisory Committee. Final report of task group II. Vol. 2 Prediction-measurement: Concepts, task analysis, principles of model construction (AFSC-TR-65-2), 1965. (AD-458 455).

Weapon System Effectiveness Industry Advisory Committee. Final report of task group II. Vol. 3 Prediction-measurement: Technical supplement (AFSC-TR-65-2). 1965. (AD-458 456).

Weinstock, G.D., Douglas, M. & Blom, B. Development of criteria and measures of effectiveness for U.S. Army tactical communications systems (Technical Report ECOM-5012-1). Paramus, NJ: Communications Systems, Inc., May 1969. (AD-881 145).

Weisz, J.D. System analysis: Manpower resources/system design integration methodology (Technical Note 9-68). Aberdeen Proving Ground, MD: Human Engineering Laboratories, August 1968. (AD-675 481).

Wellman, L.N. & Neill, W.H. The development of performance measurement standards for the USAF base level supply system (SLSR-12-72B, Master's Thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology, September 1972. (AD-750 912).

Wherry, R.J. The development of sophisticated models of man-machine system performance. In G.W. Levy (Ed.), Symposium on applied model of man-machine systems performance (NR69H-591). Columbus, OH: North American Aviation, November 1968. (AD-697 939).

Williams, G.S. A methodology to establish the criticality of attributes in operational tests (Master's Thesis). Georgia Institute of Technology, 1975. (AD-A024 199).

Williams, H.L. Dependent models for estimating human performance reliability. In R.E. Blanchard & D.H. Harris (Eds.), Man-machine effectiveness analysis. Papers presented at The Human Factors Symposium, University of California at Los Angeles, June 15, 1967. (AD-735 718).

Williams, R.L., Long, T.W. & Windholz, W.M. A standardized munitions effectiveness methodology (K-75-57 R). Colorado Springs, CO: Kaman Sciences Corporation, July 1975. (AD-B009 483).

Willis, J.E. Feasibility of the development and utilization of personnel performance effectiveness measures for man/machine function allocation decisions (Research Memorandum SRM 68-7). San Diego, CA: U.S. Naval Personnel Research Activity, October 1967. (AD-660 003).

Wortman, D.B., Hixson, A.F. & Jorgensen, C.C. A SAINT model of the AN/TSQ-73 guided missile air defense system (Research Memorandum 79). W. Lafayette, IN: Pritsker and Associates, Inc., January 1979.

Wortman, D.B., Sigal, C.E. Pritsker, A.S. & Seifert, D.J. New SAINT concepts and the SAINT II simulation program (AMRL-TR-74-119). W. Lafayette, IN: Pritsker and Associates, Inc., April 1975. (AD-A014 814).